UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas

“Una casidistancia homogénea de segundo orden para Programación cuasi-convexa”

Trabajo Especial de Grado presentado por

Br. Elvis M. Rodríguez T.

como requisito final
para obtener el título de Licenciado
en Ciencias Matemáticas
Área de Conocimiento: Optimización.
Tutor: Dr. Rómulo Castillo

Barquisimeto, Venezuela. Julio de 2011

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

ACTA
TRABAJO ESPECIAL DE GRADO

Los suscritos miembros del Jurado designado por el Jefe del Departamento de Matemáticas del Decanato de Ciencias y Tecnología de la Universidad Centroccidental "Lisandro Alvarado", para examinar y dictar el veredicto sobre el Trabajo Especial de Grado titulado:

"Una casidistancia homogénea de segundo orden para Programación cuasi-convexa"

Presentado por el ciudadano Br. Elvis M. Rodríguez T. titular de la Cédula de Identidad Nº 17.814335. Con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciado en Ciencias Matemáticas.

Luego de realizada la Defensa y en los términos que imponen los Lineamientos para el Trabajo Especial de Grado de la Licenciatura en Ciencias Matemáticas, se procedió a discutirlo con el interesado habiéndose emitido el veredicto que a continuación se expresa:

[1] ____________________

Con una calificación de ____________________ puntos.

En fe de lo expuesto firmamos la presente Acta en la Ciudad de Barquisimeto a los _____ días del mes de ____________________ de __________.

TUTOR	FIRMA
PRINCIPAL	FIRMA
PRINCIPAL	FIRMA

OBSERVACIONES:

____

____

____

[1] Aprobado ó Reprobado

Dedicado a Dios Todopoderoso, a mis Padres Irma y Rafael y a mis Hermanos Rafael, Esneyder y Nohely.

AGRADECIMIENTOS

A **Dios, todopoderoso**, por ser mi guía, mi proveedor, y darme la fortaleza para alcanzar esta meta.

A **mi madre**, serás siempre mi inspiración para lograr mis objetivos. Gracias por darme fuerzas cuando la necesité, por enseñarme que de todo se aprende, gracias por tus sabios consejos.

A **mi padre**, por todo el cariño y apoyo brindado. Por ser un excelente padre y ayudarme en todos los momentos de mi vida.

A mi **hermanos**, por su apoyo, comprensión, y porque siempre he contado con ellos en los momentos más difíciles, por orientarme siempre a lo largo de mi carrera y ser siempre ejemplos e inspiración para mi.

A **mi novia**, Rosa, por su apoyo constante e incondicional. Gracias por tu paciencia, por brindarme ese cariño tan grande todos los días desde que estás conmigo.

A mi tutor, **Rómulo Castillo**, por su asesoría y dirección en el cumplimiento de este proyecto, gracias por su tiempo, dedicación y confianza.

A **mis amigos**, Glendy, Datsy, Dannymar, Masiel, Aura, kathe, Migdalia, Joan, Beatriz, kissy, José, quienes me han acompañado durante toda mi carrera universitaria, compartiendo grandes y pequeños momentos y brindándome todo su apoyo, gracias por hacer esos días de estudios más divertidos.

A **mis mejores amigos**, Javier, El Gocho y Joelviz quienes han sido como unos hermanos para mí desde que llegue a la universidad y por compartir todos aquellos llamados "PQC" que siempre nos dieron dolores de cabeza. Gracias por su amistad sincera e incondicional.

A cada uno de los **profesores** que participaron en mi desarrollo profesional e integral, sin su ayuda y conocimientos no estaría donde me encuentro ahora, con especial agradecimiento a las profesoras Mireya Bracamonte, María Luisa Capodieci, Jurancy Ereú, Luz y a los profesores José Luis Linarez, Victor Caruci, Rómulo Castillo, Alexander Carrasco, Miguel Vivas, Mario Rodríguez, Eibar Hernández, Javier Hernández.

A **Personal de Biblioteca**, Edith Mary, Leida Rosa, Josefina, Ivom, Maritza, Yanet, Dasha, Arturo, Marlene, Claudia, Luis y Yilvric por estar siempre allí colaborando conmigo siempre que lo necesité y por darme la oportunidad de aprender y compartir el trabajo realizado en la biblioteca.

"Una casidistancia homogénea de segundo orden para Programación cuasi-convexa"

Resumen

El algoritmo de punto proximal para minimizar una función convexa $f(x)$ en $\mathbb{R}^n$ genera una sucesión $\{x^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, dada por el siguiente esquema:

$$x^{k+1} = \underset{x\geqslant 0}{\operatorname{argmín}}\{f(x) + \lambda_k\|x - x^k\|^2\}, \tag{1}$$

donde λ_k es una sucesión de números positivos y $\|.\|$ denota la norma euclideana en $\mathbb{R}^n$. Este método fue originalmente introducido por Martinet. El algoritmo de punto proximal fue posteriormente estudiado y desarrollado por Rockafellar. En adelante, se estudiaran 2 tipos de algoritmos de puntos proximales no cuadráticos pero reemplazando la distancia cuadrática (1) por una distancia de Bregman o casi distancia entrópica. Entre los tipos de casi distancia entrópicas está la casidistancia inducida por la $\varphi - divergencia$ y viene expresada como:

$$d_\varphi(x, y) = \sum_{i=1}^{n} y_i\varphi(x_i/y_i), \tag{2}$$

donde $\varphi : \mathbb{R} \to (-\infty, +\infty)$ es una función propia y estrictamente convexa que satisface ciertas propiedades.

Una opción de φ es el caso en que $\varphi(t) = t\ln(t) - t + 1$, la cual, con la correspondiente d_φ es conocida como la función Kullback-Leibler entrópica, para estadística y terminología entrópica.

El algoritmo proximal basando en φ-divergencia esta originalmente designado para una función convexa sujeto a constantes no negativas $x \geqslant 0$, la cual genera una sucesión definida como:

$$\begin{cases} x^0 > 0 \\ x^{k+1} = \underset{x\geqslant 0}{\operatorname{argmín}}\{f(x) + \lambda_k d_\varphi(x, x^k)\}. \end{cases} \tag{3}$$

La intención es enfocarnos, en clases de algoritmos proximales de la forma (3) pero con una casidistancia homogénea de segundo orden, definida por:

$$d_\phi(x,y) = \sum_{i=1}^{n} y_i^2 \phi(x_i/y_i), \tag{4}$$

donde ϕ se define como el Kernel y está definida con 2 tipos de φ y una función cuadrática, es decir:

$$\phi(t) = \mu\varphi(t) + \frac{\nu}{2}(t-1)^2, \tag{5}$$

donde los parámetros $\mu > 0$ y $\nu \geqslant 0$.

Esta función ϕ satisface ciertas propiedades que son desarrolladas para la minimización y la convergencia del siguiente problema:

$$\begin{aligned} \text{mín} \quad & f(x) \\ s.a. \quad & x \geqslant 0, \end{aligned} \tag{6}$$

en el cual, $f : \mathbb{R}^n \to \mathbb{R}$ es una función propia, cuasi-convexa, semicontinua inferior. Así, es claro que no necesitamos la convexidad de la función y además la sucesión que se genera es de la forma:

$$\begin{cases} x^0 > 0 \\ x^{k+1} = \underset{x \geqslant 0}{\text{argmín}} \;\; \{f(x) + \lambda_k d_\phi(x, x^k)\}, \end{cases} \tag{7}$$

donde λ_k es dada como en (1). El objetivo principal de este trabajo es establecer la convergencia global de la sucesión generada por (7) y bajo ciertas hipótesis resolver el problema (6) definido anteriormente.

Índice

Capítulo 1

Introducción.

A continuación presentaremos un poco el desarrollo del paper de Shaohua Pan y Jein-Shan Chen, publicado en el año 2007. La intención es hacer una regularización, la cual surge principalmete de los problemas mal puestos. Dado el problema de la forma:

$$L : \mathbb{X} \longrightarrow \mathbb{X}$$
$$L(f) = 0, \, donde \; f \in \mathbb{X} \, (Espacio \; de \; funciones)$$

y L es el operador (usualmente diferencial o integro diferencial).

Este problema se dice que está mal puesto, cuando no tiene solución o tiene más de una solución o tiene solución única, pero la solución no depende de manera continua de algunos parámetros del operador L.

La idea de este concepto, es reemplazar el problema inicial por el operador: $L + \lambda M$, donde $\lambda \in \mathbb{R}$ y $M : \mathbb{X} \to \mathbb{X}$ es tal que el problema:

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

esté bien puesto para algún $\lambda > 0$.

En adelante, se utilizará una regularización de este tipo y se probará bajo ciertas hipótesis la convergencia de dos métodos de puntos proximales con casidistancias para resolver un determinado problema primal.

Capítulo 2

Preliminares

En este capítulo se definirán algunos conceptos básicos para ser ultilizados en el desarrollo del trabajo, asi como la convergencia de un par de métodos de punto proximal, uno con la distancia euclideana y otro con Distancias de Bregman, los cuales nos orientarán hacia la convergencia de los 2 métodos principales de este trabajo.

§2.1. El algoritmo de punto proximal para optimización en $\mathbb{R}^n$ con la distancia euclideana

Definición 2.1. *Una sucesión $\{y^k\}_{k\in N}$ es Fejér convergente en un conjunto no vacío $\mathbb{U} \subseteq \mathbb{R}^n$, con respecto a una función casi distancia $d(\cdot,\cdot)$, si para cada $u \in U$, tenemos que $d(u, y^{k+1}) \leqslant d(u, y^k)$. Cuando d es la distancia euclídea, $\{y^k\}_{k\in\mathbb{N}}$ se llama Fejér convergente en $\mathbb{U}$ y en este caso se tiene:*

$$\|y^{k+1} - u\| \leqslant \|y^k - u\|, \quad \forall k \geqslant 0. \tag{2.1}$$

Definición 2.2. *Dada una función real extendida $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, denotaremos su dominio por:*

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

y su epígrafo como:

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

Además, f se dice que es propia si $domf \neq \emptyset$ y $f(x) > -\infty$ para cualquier $x \in domf$. f se dirá que es una función semicontinua inferior si $epif$ es un subconjunto cerrado de $\mathbb{R}^n \times \mathbb{R}$.

Definición 2.3. *Una función $f : R^n \to R$ se dice que es coerciva si y solo si*

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

Proposición 2.1. *Si $\{y^k\}_{k\in\mathbb{N}}$ es Fejér convergente en $\mathbb{U} \neq \emptyset$ entonces, $\{y^k\}_{k\in\mathbb{N}}$ está acotada. Si y es punto de acumulación de $\{y^k\}_{k\in\mathbb{N}}$ a lo largo de $\mathbb{U}$ entonces,*

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

Demostración. Como $\{y^k\}_{k\in\mathbb{N}}$ es Fejér convergente, por la defínción 2.1 se tiene que:

$$\|y^{k+1} - u\| \leqslant \|y^k - u\|, \forall\ k \geqslant 0, \forall\ u \in \mathbb{U},$$

esto implica que:

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

Luego, se tiene que $\{y^k\}_{k\in\mathbb{N}}$ está contenida en una bola de centro u y radio $\|y^0 - u\|$, así $\{y^k\}_{k\in\mathbb{N}}$ está acotada. Para la segunda parte de la pueba, como y es punto de acumulación entonces existe una subsucesión $\{y^{f_k}\}_{f_k\in\mathbb{N}}$ de $\{y^k\}_{k\in\mathbb{N}}$ que converge a y. Por otra parte, debido a que $\{y^k\}_{k\in\mathbb{N}}$ es Fejér convergente y como y está en $\mathbb{U}$ entonces $\{\|y^k - y\|\}_{k\in\mathbb{N}}$ es decreciente y no negativa, asi se tiene que la subsucesión $\{\|y^{f_k} - y\|\}_{f_k\in\mathbb{N}}$ converge a cero. Entonces el conjunto de subsucesiones convergen a cero. Así, $\lim_{k\to\infty} \|y^k - y\| = 0$ implicando que $\lim_{k\to\infty} y^k = y$. ∎

Ahora, definamos el algoritmo de punto proximal y su convergencia. El problema de interés es el siguiente:

$$\begin{aligned} \textbf{min}\ \ & f(x) \\ \textbf{s.a}\ \ & x \in \overline{S}, \end{aligned} \tag{2.2}$$

donde f es una función convexa y continuamente diferenciable. Ahora se define, la convergencia del algoritmo de minimización con el siguiente teorema.

Teorema 2.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ convexa y continuamente diferenciable. Supongamos que el conjunto $\mathbb{U}$ de minimizadores de f en $\mathbb{R}^n$ es no vacío. Entonces la sucesión $\{x^k\}_{k\in\mathbb{N}}$ dada por:*

$$x^0 \in \mathbb{R}^n$$
$$x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \{f(x) + \lambda_k \|x - x^k\|^2\}, \tag{2.3}$$

donde λ_k son números reales que satisfacen:

$$0 < \lambda_k \leqslant \bar{\lambda} \quad \forall k \in \mathbb{N}, \tag{2.4}$$

para algún $\bar{\lambda} > 0$. Converge a un punto $x^* \in \mathbb{U}$.

Demostración. Para el desarrollo de la demostración vamos a considerar 4 pasos. En el primer paso, se probará que la sucesión definida en (2.3) está bien definida. En segunda instancia veremos que $\{x^k\}_{k\in\mathbb{N}}$ es Fejér convergente en $\mathbb{U}$. En el paso 3 se definirá el factor $\lim_{k\to\infty}(x^{k+1} - x^k) = 0$, el cual se usará en el paso 4, donde se probará que el conjunto de puntos de la sucesión $\{x^k\}_{k\in\mathbb{N}}$ pertenecen a $\mathbb{U}$. Finalmente, los pasos desde el 2 hasta el 4 y conjuntamente con la proposición (2.1) se obtiene la prueba del teorema.

Paso 1 La sucesión $\{x^k\}_{k\in\mathbb{N}}$ está bien definida.

Procedamos por inducción. Para $k = 0$, se satisface pues simplemente se toma un punto en $\mathbb{R}^n$. Para los términos $k+1$, sea $f_k(x) = f(x) + \lambda_k\|x - x^k\|^2$. Luego, por hipótesis f alcanza un mínimo, ya que f está acotada inferiormente, además el término cuadrático $\lambda_k\|x - x^k\|^2$ nos dice que $\lim_{\|x\|\to+\infty} f_k(x) = +\infty$; por continuidad de f, f_k es contínua y además por la minimización en (2.2) reduce el problema a un conjunto compacto, garantizándose así solución a f_k. Por otra parte como f es convexa y el término $\lambda_k\|x - x^k\|^2$ es estrictamente convexo entonces f_k será una función estrictamente convexa para algún λ_k y de esta manera se garantizará que el mínimo generado en cada término de la sucesión $\{x^{k+1}\}_{k\in\mathbb{N}}$ es único.

Paso 2. En este caso se probará que:

$$\|x^{k+1} - \bar{x}\|^2 \leqslant \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2, \ \forall k \geqslant 0 \ \textit{y todo} \ \bar{x} \in \mathbb{U}. \tag{2.5}$$

Para ello:

$$\begin{aligned}\|x^k - \bar{x}\|^2 &= \|x^k - x^{k+1} + x^{k+1} - \bar{x}\|^2 \\ &= \|x^k - x^{k+1}\|^2 + \|x^{k+1} - \bar{x}\|^2 + 2\langle x^k - x^{k+1}, x^{k+1} - \bar{x}\rangle. \end{aligned} \tag{2.6}$$

Luego, como x^{k+1} resuelve (2.3) se tiene:

$$
\begin{aligned}
& 0 = \nabla f_k(x^{k+1}) = \nabla f(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k) \\
\Rightarrow \; & 2(x^{k+1} - x^k) = -\frac{1}{\lambda_k}\nabla f(x^{k+1}),
\end{aligned} \tag{2.7}
$$

así, de (2.6) y (2.7) se tiene que:

$$
\begin{aligned}
\|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2 - \|x^{k+1} - \bar{x}\|^2 &= 2\langle x^k - x^{k+1}, x^{k+1} - \bar{x}\rangle \\
&= \frac{1}{\lambda_k}\langle \nabla f(x^{k+1}), x^{k+1} - \bar{x}\rangle \\
&\geqslant \frac{1}{\lambda_k}[f(x^{k+1}) - f(\bar{x})] \geqslant 0,
\end{aligned} \tag{2.8}
$$

Puesto que $\bar{x}$ es el minimizador de f. Por tanto, reagrupando los términos en la última desigualdad (2.8) obtenemos lo deseado.

Paso 3 Veamos ahora que:

$$
\lim_{k\to\infty}(x^{k+1} - x^k) = 0. \tag{2.9}
$$

A partir de la conclusión del paso 2 se puede obtener:

$$
\begin{aligned}
\|x^{k+1} - \bar{x}\|^2 &\leqslant \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2 \\
&\leqslant \|x^k - \bar{x}\|^2.
\end{aligned}
$$

Asi, $\{\|x^k - \bar{x}\|\}_{k\in\mathbb{N}}$ es decreciente y acotada pues:

$$
0 \leqslant \|x^{k+1} - \bar{x}\| \leqslant \|x^k - \bar{x}\| \leqslant \cdots \leqslant \|x^0 - \bar{x}\|.
$$

Por tanto, la sucesión converge, es decir $\{\|x^k - \bar{x}\|\} \to 0$, con lo cual se tiene (2.9).

Paso 4. $\{x^k\}_{k\in\mathbb{N}}$ tiene puntos de acumulación y todos ellos pertenecen a $\mathbb{U}$.

La existencia de los puntos de acumulación estarán garantizados por el paso 2, pues en ese caso se obtuvo la fejér convergencia de $\{x^k\}_{k\in\mathbb{N}}$ y usando esa condición en la proposicion (2.1) tenemos que $\{x^k\}_{k\in\mathbb{N}}$ es acotada, quedando asi garantizada los puntos de acumulación sucesión. Sea $\bar{x}$ un punto de acumulación de la sucesión $\{x^k\}_{k\in\mathbb{N}}$ y tomemos $\{x^{f_k}\}_{k\in\mathbb{N}}$ una subsucesión de $\{x^k\}_{k\in\mathbb{N}}$, tal que $\lim_{k\to\infty} x^{f_k} = \bar{x}$. Por (2.7):

$$\nabla f(x^{f_k+1}) = 2\lambda_{j_k}(x^{f_k} - x^{f_k+1}). \tag{2.10}$$

Luego, por el paso 3, $\lim_{fk\to\infty} x^{f_k+1} = \lim_{fk\to\infty} x^{f_k} = \bar{x}$. Asi, tomando límite cuando $fk \to \infty$ en la igualdad (2.10), usando $\lambda_{jk} \leqslant \bar{\lambda}$ y además que f es continuamente diferenciable, tendríamos $\nabla f(\bar{x}) = 0$. Por tanto, por la convexidad de f, $\bar{x} \in \mathbb{U}$.

De esta forma por los casos 2-4 y la proposición (2.1), existe $x^* \in \mathbb{U}$ tal que

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

$\blacksquare$

DEFINICIÓN 2.4. *ξ es un subgradiente de f en x si y solo si*

$$\langle \xi, y - x\rangle \leqslant f(y) - f(x), \quad \forall y \in \mathbb{R}^n. \tag{2.11}$$

El conjunto de subgradiente $\partial f(x) = \{\xi : \xi$ es un subgradiente de f en $x\}$. se define como el subdiferencial de f en x.

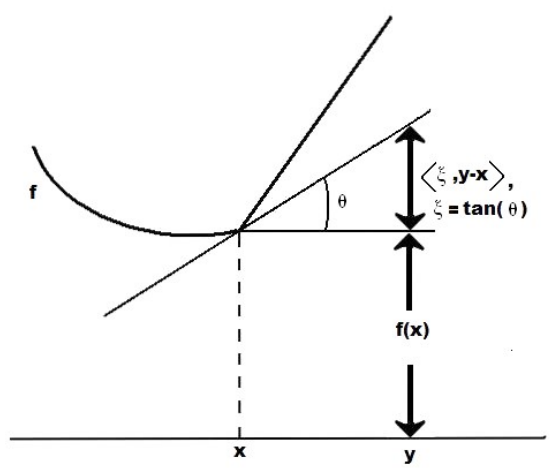

FIGURA 2.1: SUBGRADIENTE

§2.2. Funciones y Distancias Bregman

Definición 2.5. *Sea S un subconjunto abierto y convexo de $\mathbb{R}^n$ y $\bar{S}$ su clausura. Considere una función h convexa de valores reales definida en $\bar{S}$ y sea $D_h : \bar{S} \times S \to \mathbb{R}$ dada por:*

$$D_h(x, y) = h(x) - h(y) - \nabla h(y)^t (x - y). \tag{2.12}$$

h es llamada función Bregman, si satisface las siguientes condiciones:

1. *h es continuamente diferenciable en S.*

2. *h es estrictamente convexa y continua en $\bar{S}$.*

3. *Para todo $\delta \in \mathbb{R}$ el conjunto de niveles parciales $\Gamma_1(y, \delta) = \{x \in \bar{S} : D_h(x, y) \leqslant \delta\}$, $\Gamma_2(x, \delta) = \{y \in S : D_h(x, y) \leqslant \delta\}$ son acotados para todo $y \in S$ y todo $x \in \bar{S}$, respectivamente.*

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

5. *Si $\{x^k\} \subset \bar{S}$ y $\{y^k\} \subset S$ son sucesiones tales que $\{x^k\}_{k \in \mathbb{N}}$ está acotada, $\lim_{k \to \infty} y^k = y^*$ y $\lim_{k \to \infty} D_h(x^k, y^k) = 0$ entonces $\lim_{k \to \infty} x^k = y^*$.*

Donde D_h es dada como en (2.12) y es llamada distancia de Bregman inducida por h y además a S se le conoce como la zona de h.

Observaciones 2.1.

1. *$D_h \geqslant 0$ para todo $x \in \bar{S}$ y $y \in S$.*
   Para tal afirmación, es suficiente ver que h es estrictamente convexa y continuamente diferenciable en S, pues de esta manera, satisface que:

$$h(x) \geqslant h(y) + \nabla h(y)^t (x - y), \ \ \forall x. \tag{2.13}$$

2. *$D_h(x, y) = 0$ si y solo si x=y.*
   ($\Longrightarrow$) $D_h(x, y) = 0$ entonces por definición:
   $h(x) - h(y) = (\nabla h(y))^t (x - y)$ Asi, usando la diferenciabilidad y convexidad de h se tiene que y=x.

   ($\Longleftarrow$) Sea x=y entonces:
   $D_h(x, y) = D_h(x, x) = h(x) - h(x) - \nabla h(x)^t (x - x) = 0 - \nabla h(x)^t (0) = 0.$

DEFINICIÓN 2.6. *Vamos a introducir dos subclases para ser usadas más adelante.*

1. *Una función Bregman es llamada coerciva en la frontera, si* $\{y^k\} \subset S$ *es tal que* $\lim\limits_{k\to\infty} \nabla h(y^k)^t(x-y^k) = -\infty$ *para todo* $x \in S$.
2. *Una función Bregman es llamada de Zona coerciva, si para cada* $y \in \mathbb{R}^n$ *entonces existe un* $x \in S$ *tal que* $\nabla h(x) = y$.

Veamos algunos ejemplos de funciones Bregman.

EJEMPLO 2.1. $S = \mathbb{R}^n$, $h(x) = x^tMx$, *con* $M \in \mathbb{R}^{n\times n}$ *simétrica y definida positiva. En este caso,* $D_h(x,y) = (x-y)^tM(x-y) = \|x-y\|_M^2$.

Solución. h es una función cuadrática, luego es continuamente diferenciable en $\mathbb{R}^n$ y además por ser M definida positiva, nos dice que la función es estrictamente convexa en $\mathbb{R}^n$. Calculemos la distancia D_h, por definición de h:

$$\begin{aligned}
D_h(x,y) &= x^tMx - y^tMy - (\nabla(y^tMy))^t(x-y)\\
&= x^tMx - y^tMy - (2My)^t(x-y)\\
&= x^tMx - y^tMy - (2y^tM^t)(x-y)\\
&= x^tMx - y^tM^ty - 2y^tM^tx + 2y^tM^ty\\
&= x^tMx - 2y^tM^tx + y^tM^ty\\
&= x^tM^tx - y^tM^tx + y^tM^ty - y^tM^tx\\
&= (x^t - y^t)M^tx + y^tM^t(y-x)\\
&= (x^t - y^t)M^tx + y^tM^t(y-x)\\
&= (x-y)^tM^tx - (x-y)^tM^ty\\
&= (x-y)^tM^t(x-y)\\
&= (x-y)^tM(x-y).
\end{aligned}$$

$\therefore D_h(x,y) = (x-y)^tM(x-y).$ ■

EJEMPLO 2.2. $S = \mathbb{R}^n_{++}$, $h(x) = \sum\limits_{j=1}^{n} x_j \log(x_j)$ *extendido con continuidad en la frontera y usando la convención de que* $0\log(0) = 0$. *Luego,* $D_h(x,y) = \sum\limits_{j=1}^{n}(x_j \log(\frac{x_j}{y_j}) + y_j - x_j)$. *Esta función es la llamada Kullback-Leiber.*

Solución. Sean $x, y \in \mathbb{R}^n_{++}$ Luego:

$$
\begin{aligned}
D_h(x,y) &= \sum_{i=1}^{n} x_i \log(x_i) - y_i \log(y_i) - (1+\log(y_i))(x_i - y_i)) \\
&= \sum_{i=1}^{n} x_i \log(x_i) - y_i \log(y_i) - x_i + y_i - x_i \log(y_i) - y_i \log(y_i) \\
&= \sum_{i=1}^{n} x_i \log(x_i) - x_i \log(y_i) + y_i - x_i \\
&= \sum_{i=1}^{n} x_i \log(x_i/y_i) + y_i - x_i.
\end{aligned}
$$

■

PROPOSICIÓN 2.2. *Si h es una función bregman con zona S entonces:*

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

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

iii) $D_h(.,y)$ *es estrictamente convexo para* $y \in S$.

Demostración.

i) Por definición tenemos:

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

$\therefore$ (i) es satisfecha.

(ii) Basta aplicar gradiente respecto de x a la definición (2.12).

(iii) Sea y fijo pero arbitrario en S y consideremos la función de una variable $f(x) = D(.,y) = h(x) - h(y) - \nabla h(y)^t(x-y)$ y veamos que para todo x, f es convexa. Tomemos $x, \bar{x} \in \bar{S}$ y la combinación $\alpha x + (1-\alpha)\bar{x}$, con $\alpha \in (0,1)$.

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

$\therefore f$ es estrictamente convexa y asi $D_h(.,y)$ es estrictamente convexa,$\forall y \in S$.

§2.3. El método de punto proximal para distancias Bregman

El problema de interés es:

$$\begin{aligned}
\min f(x) \\
s.a. \;\; x \in \overline{S},
\end{aligned} \tag{2.14}$$

donde $S \subset \mathbb{R}^n$ es abierto y convexo, $\bar{S}$ es la clausura de S y f es una función contínua y convexa en $\bar{S}$. El método de punto proximal con distancias Bregman está definido como:

$$\begin{aligned}
x^0 &\in S \\
x^{k+1} &= arg \min_{x \in \bar{S}} \{f(x) + \lambda_k D_h(x, x^k)\},
\end{aligned} \tag{2.15}$$

donde h es una función Bregman con Zona S y λ_k satisface:

$$0 < \lambda_k \leqslant \bar{\lambda}, \tag{2.16}$$

para algún $\bar{\lambda} > 0$. Estudiemos ahora la convergencia de dicha sucesión. Para esto, asumiremos que (2.14) tiene solución y además que f es acotada inferiormente en $\overline{S}$.

Teorema 2.2. *Si el problema (2.14) tiene solución y h es coerciva en la frontera con respecto al conjunto S entonces la sucesión $\{x^k\}_{k\in\mathbb{N}}$ generada por (2.15) converge a una solución x^* del problema (2.14) pero usando la distancia de Bregman.*

Demostración. Para la demostración de este teorema, se procederá de manera análoga a la demostración del teorema de punto proximal para $\mathbb{R}^n$

Paso 1 La sucesión esta bien definida y contenida en S.

Sea β una solución para el problema (2.14) en $\bar{S}$ y sea $f_k(x) = f(x) + \lambda_k D_k(x, x^k)$. Entonces:

$$\begin{aligned}\beta \leqslant f(x) &\Longleftrightarrow \beta + \lambda_k D_k(x, x^k) \leqslant f(x) + \lambda_k D_k(x, x^k)\\ &\Longleftrightarrow \beta + \lambda_k D_k(x, x^k) \leqslant f_k(x).\end{aligned}$$

Definamos ahora el conjunto de niveles de f_k por:

$$L_{f_k}(\gamma) = \{x \in \mathbb{R}^n \,:\, f_k(x) \leqslant \gamma\}. \tag{2.17}$$

Ahora veamos que $L_{f_k} \subset \Gamma_1$ donde Γ_1 es como en la definición 2.5 (3).

$$\begin{aligned}x \in L_{f_k} &\Longrightarrow f_k(x) \leqslant \gamma\\ &\Longrightarrow f(x) + \lambda_k D_k(x, x^k) \leqslant \gamma\\ &\Longrightarrow \beta + \lambda_k D_k(x, x^k) \leqslant \gamma\\ &\Longrightarrow \lambda_k D_k(x, x^k) \leqslant \gamma - \beta\\ &\Longrightarrow D_k(x, x^k) \leqslant (\gamma - \beta)(\lambda_k)^{-1}\\ &\Longrightarrow x \in \Gamma_1.\end{aligned}$$

$\therefore\ L_{f_k} \subset \Gamma_1$

Asi, usando la definición 2.5 parte (3) se sigue que el conjunto de niveles de f_k es acotado, luego la minimización en (2.15) reduce el problema a un conjunto compacto y asi el mínimo está logrado. Por otro lado, f_k es estrictamente convexa, pues $D_k(\cdot, x^k)$ es estrictamente convexa por proposición 2.2 (iii), asi el mínimo es único y $\{x^k\}_{k\in\mathbb{N}}$ estará bien definida.

Veamos ahora que la sucesión $\{x^{k+1}\}_{k\in\mathbb{N}} \in S$. Es fácil ver para (2.15) que $\{x^{k+1}\}_{k\in\mathbb{N}}$ es el único $x \in \bar{S}$ tal que:

$$\lambda_k \nabla h(x^k) \in \partial(f + \lambda_k h)(x). \tag{2.18}$$

Se probará bajo la definición 2.6 (1) que $\partial(f+\lambda_k h)(x) = \emptyset$ para todo $x \in \partial(S)$ lo cual implica en vista de (2.15), que $\{x^{k+1}\}_{k\in\mathbb{N}} \in S$. Tomemos $x \in \partial S$ y supongamos que existe $\xi \in \partial(f + \lambda_k h)(x)$, además $z \in S$ y definamos:

$$y^\ell = (1-\varepsilon_\ell)x + \varepsilon_\ell z, \; donde \; \lim_{\ell\to\infty} \varepsilon_\ell = 0 \; y \; \varepsilon_\ell \neq 1. \tag{2.19}$$

Trabajando (2.19), podemos obtener:

$$y^\ell = (1-\varepsilon_\ell)x + \varepsilon_\ell z \Longleftrightarrow y^\ell - x = \varepsilon_\ell(z - x). \tag{2.20}$$

Por otro lado, por la convexidad de f en $\bar{S}$ tenemos que:

$$\begin{aligned} f(y^\ell) - f(x) &= f((1-\varepsilon_\ell)x + \varepsilon_\ell z) - f(x) \\ &\leqslant (1-\varepsilon_\ell)f(x) + \varepsilon_\ell f(z) - f(x) \\ &= \varepsilon_\ell(f(z) - f(x)) \end{aligned} \tag{2.21}$$

y además:

$$\begin{aligned} y^\ell = (1-\varepsilon_\ell)x + \varepsilon_\ell z &\Leftrightarrow y^\ell = x - \varepsilon_\ell x + \varepsilon_\ell z \\ &\Leftrightarrow y^\ell - x + x\varepsilon_\ell = \varepsilon_\ell z \\ &\Leftrightarrow y^\ell - \varepsilon_\ell y^\ell - x + x\varepsilon_\ell = \varepsilon_\ell z - \varepsilon_\ell y^\ell \\ &\Leftrightarrow (y^\ell - x)(1-\varepsilon_\ell) = \varepsilon_\ell(z - y^\ell) \\ &\Leftrightarrow (y^\ell - x) = \frac{\varepsilon_\ell}{1-\varepsilon_\ell}(z - y^\ell). \end{aligned} \tag{2.22}$$

Luego por la definición (2.19) tenemos que $y^\ell \in S$, por la convexidad de S y $\lim_{\ell\to\infty} y^\ell = x$. Además:

$$\begin{aligned}
\varepsilon_\ell \xi^t(z-x) &= \xi^t(y^\ell - x) && (Por\ 2{,}20)\\
&\leqslant f(y^\ell) - f(x) + \lambda_k(h(y^\ell) - h(x)) && (definicion\ 2{,}4)\\
&\leqslant f(y^\ell) - f(x) + \lambda_k \nabla h(y^\ell)^t(y^\ell - x) && (diferenciabilidad\ h)\\
&\leqslant \varepsilon_\ell(f(z) - f(x)) + \lambda_k \frac{\varepsilon_\ell}{1-\varepsilon_\ell}\nabla h(y^\ell)^t(z - y^\ell). && (por\ 2{,}21\ y\ 2{,}22)
\end{aligned}$$

Esta última desigualdad, agrupando los términos obtenemos:

$$\begin{aligned}
& \varepsilon_\ell \xi^t(z-x) \leqslant \varepsilon_\ell(f(z) - f(x)) + \lambda_k \frac{\varepsilon_\ell}{1-\varepsilon_\ell}\nabla h(y^\ell)^t(z-y^\ell)\\
\Leftrightarrow\ & \varepsilon_\ell(\xi^t(z-x) - (f(z) - f(x))) \leqslant \lambda_k \frac{\varepsilon_\ell}{1-\varepsilon_\ell}\nabla h(y^\ell)^t(z-y^\ell)\\
\Leftrightarrow\ & (\xi^t(z-x) - (f(z) - f(x))) \leqslant \lambda_k \frac{1}{1-\varepsilon_\ell}\nabla h(y^\ell)^t(z-y^\ell)\\
\Leftrightarrow\ & \frac{1-\varepsilon_\ell}{\lambda_k}((f(x) - f(z)) + \xi^t(z-x)) \leqslant \nabla h(y^\ell)^t(z-y^\ell).
\end{aligned}$$

Ahora bien, como $\lim_{\ell\to\infty} y^\ell = x \in \partial S$, por definición 2.6(1) implica que la parte derecha de la última desigualdad tiende a $-\infty$,cuando $\ell \to \infty$ por definición; pero, la parte izquierda de la misma tiende a un número finito, por tanto es una contradicción. Asi, $\partial(f + \lambda_k h) = \emptyset, \forall x \in \partial S$ y con esto se concluye que $\{x^{k+1}\}_{k\in\mathbb{N}} \in S$

Paso 2 Veamos que $D_h(\bar{x}, x^{k+1}) \leqslant D_h(\bar{x}, x^k) - D_h(x^{k+1}, x^k)$ para todo k y cada solución $\bar{x}$ de (2.14).

Usando la proposición 2.2(i), con $x = \bar{x}, y = x^k, z = x^{k+1}$ se tiene que:

$$D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^{k+1} - \bar{x}\rangle. \quad (2.23)$$

Por (2.15) como x^{k+1} es solución se tiene:

$$0 \in \partial[f + \lambda D_h(\cdot, x^k)](x^{k+1}). \quad (2.24)$$

Ahora, por (2.24) y proposición 2.2(ii):

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

Sea $y^k = \lambda_k(\nabla h(x^k) - \nabla h(x^{k+1}))$. Usando, nuevamente (2.23) y la definición de subgradiente (2.4) tenemos:

$$
\begin{aligned}
D_h(\bar{x},x^k)-D_h(\bar{x},x^{k+1})-D_h(x^{k+1},x^k) =& \langle\frac{1}{\lambda_k}(\lambda_k\nabla h(x^k)-\nabla h(x^{k+1})),x^{k+1}-\bar{x}\rangle\\
=&\frac{1}{\lambda_k}\langle y^k,x^{k+1}-\bar{x}\rangle\\
\geqslant&\frac{1}{\lambda_k}(f(x^{k+1})-f(\bar{x})) \qquad (2.25)\\
\geqslant&0 \qquad (\bar{x}\, es\, minimizador\, de\, f\, en\, \overline{S})
\end{aligned}
$$

$$\therefore\; D_h(\bar{x},x^k)-D_h(\bar{x},x^{k+1})-D_h(x^{k+1},x^k)\geqslant 0$$

así:

$$D_h(\bar{x},x^{k+1})\leqslant D_h(\bar{x},x^k)-D_h(x^{k+1},x^k) \tag{2.26}$$

Paso 3 La sucesión $\{x^k\}_{k\in\mathbb{N}}$ está acotada y $\lim\limits_{k\to\infty} x^{f_k}=\widehat{x}$ implica que $\lim\limits_{k\to\infty} x^{f_k+1}=\widehat{x}$.

Por el paso 2, se puede ver que $\{D_k(\bar{x},x^k)\}$ es decreciente y además por la no negatividad de $D_k(x,y)$ se obtiene que la sucesión $\{D_h(\bar{x},x^k)\}$ está acotada y en consecuencia es convergente. Ahora, en (2.26), aplicando límite a ambos lados de la desigualdad y el hecho de que h es continua se tiene:

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

Luego, por la definición 2.5(3) $\{x^k\}_{k\in\mathbb{N}}$ está acotada. Ahora si consideramos, $\lim\limits_{k\to\infty} x^{f_k}=\widehat{x}$ donde $\{x^{f_k}\}$ es una subsucesión de x^k, y además lo obtenido en (2.27) se satisfacen las hipótesis de la definición 2.5(5) y de esta manera se puede concluir que:

$$\lim_{k\to\infty} x^{f_k+1}=\widehat{x}.$$

Paso 4. La sucesión $\{x^k\}$ tiene todos los puntos frontera y cada uno son soluciones de (2.14).

Tomemos un punto $\bar{x}$ solución de (2.14). Sea $\widehat{x}$ un punto en la clausura de $\{x^k\}_{k\in\mathbb{N}}$ y $\{x^{f_k}\}$ una subsucesión de $\{x^k\}_{k\in\mathbb{N}}$, tal que $\lim\limits_{k\to\infty} x^{f_k}=\widehat{x}$. La existencia de $\widehat{x}$ sigue del *Caso* 3 donde también aseguramos que $\lim\limits_{k\to\infty} x^{f_k+1}=\widehat{x}$. Ahora, por (2.25)

$$
\begin{aligned}
0 &\leqslant \tfrac{1}{\lambda}(f(x^{f_k+1})-f(\bar{x}))\\
&\leqslant \tfrac{1}{\lambda_k}(f(x^{f_k+1})-f(\bar{x}))\\
&\leqslant D_h(\bar{x},x^{f_k})-D_h(\bar{x},x^{f_k+1})-D_h(x^{f_k+1},x^{f_k}) \xrightarrow{k\to\infty} 0\,,
\end{aligned} \tag{2.28}
$$

Lo cual se obtiene por la convergencia de $D_h(\bar{x}, x^k)$ y (2.27). Luego, tomando límite cuando $k \to \infty$ en (2.28), se concluye que $f(\widehat{x}) = f(\bar{x})$. Asi, como $\overline{S}$ es cerrado y $\{x^k\} \subset \overline{S}$ se tiene $\widehat{x}$ resuelve (2.14).

Para terminar la prueba necesitamos el teorema de la convergencia Fejér para distancias Bregman, lo cual es cierto, pero se procederá de manera directa. Sea $\widehat{x}$ un punto clausura de $\{x^k\}_{k\in\mathbb{N}}$ y consideremos una subsucesión $\{x^{f_k}\}$ de $\{x^k\}_{k\in\mathbb{N}}$ tal que $\lim\limits_{k\to\infty} x^{f_k} = \widehat{x}$. Por la definición 2.5(4) $\lim\limits_{k\to\infty} D_h(x^{f_k}, \widehat{x}) = 0$; Pero por el paso 4, $\widehat{x}$ resuelve el problema (2.14) y por los pasos 2 y 3 $\{D_h(\widehat{x}, x^k)\}$ es una sucesión no negativa, decreciente y con una subsucesión convergiendo a cero. De acá sigue que toda sucesión converge a cero y por definición 2.5(4) se tiene $\lim\limits_{k\to\infty} x^k = \widehat{x}$. ∎

DEFINICIÓN 2.7. *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función propia semicontinua inferior. Para cada $x \in domf$, el subdiferencial de Fréchet para f en x, denotado por $\widehat{\partial} f(x)$, es el conjunto de vectores $s \in \mathbb{R}^n$ tal que:*

$$\liminf_{y\to x\; y\neq x} \frac{f(y) - f(x) - \langle s, x - y\rangle}{\|y - x\|} \geqslant 0. \tag{2.29}$$

Si $x \notin domf$, entonces $\widehat{\partial} f(x) = \emptyset$. El vector s que satisface la desigualdad (2.29) también es llamado subgradiente regular de f en x.

No es difícil ver que la desigualdad (2.29) es equivalente a $f(y) \geqslant f(x) + \langle s, y - x\rangle + o(\|y - x\|)$,

donde:

$$\lim_{y\to x} \frac{o(\|y - x\|)}{\|y - x\|} = 0.$$

Pues:

$$\begin{aligned}
&& f(y) \geqslant & f(x) + \langle s, y-x\rangle + \circ(\|y-x\|) \\
\Rightarrow && f(y) - f(x) - \langle s, y-x\rangle \geqslant & \circ(\|y-x\|) \\
\Rightarrow && \frac{f(y) - f(x) - \langle s, y-x\rangle}{\|y-x\|} \geqslant & \frac{\circ(\|y-x\|)}{\|y-x\|}, \ \forall\, y \neq x \\
\Rightarrow && \liminf_{y\to x\ y\neq x} \frac{f(y) - f(x) - \langle s, y-x\rangle}{\|y-x\|} \geqslant & \liminf_{y\to x\ y\neq x} \frac{\circ(\|y-x\|)}{\|y-x\|} \\
\Rightarrow && \liminf_{y\to x\ y\neq x} \frac{f(y) - f(x) - \langle s, y-x\rangle}{\|y-x\|} \geqslant & 0
\end{aligned}$$

DEFINICIÓN 2.8. *Suponga que $f : \mathbb{R}^n \to \mathbb{R}$ es una función diferenciable. Entonces:*

(a) *Para el problema de optimización sin restricciones de minimizar una función $f(x)$ sobre $x \in \mathbb{R}^n$,x^* es llamado punto estacionario, si $\nabla f(x^*) = 0$.*

(b) *Para un problema de optimizacón con restriciones de minimizar una función $f(x)$ sobre C, donde C es un subconjunto no vacío de $\mathbb{R}^n$, x^* es llamado punto estacionario, si:*

$$\nabla f(x^*)^T(x - x^*) \geqslant 0, \ \forall x \in C.$$

DEFINICIÓN 2.9. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función propia. Entonces, f es llamada cuasi-convexa, si para todo $x, y \in domf$ y $\beta \in (0,1)$ se cumple que:*

$$f(\beta x + (1-\beta)y) \leqslant max\{f(x), f(y)\}.$$

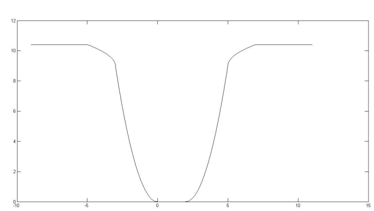

FIGURA 2.2: CUASI-CONVEXA

DEFINICIÓN 2.10. *sea $f : \mathbb{R}^n \to \mathbb{R}$ una función propia. Entonces, f es llamada estrictamente cuasi-convexa, si $\forall x, y \in domf$ con $f(x) \neq f(y)$ se cumple que:*

$$f(\beta x + (1-\beta)y) < max\{f(x), f(y)\}\ \forall \beta \in (0,1). \tag{2.30}$$

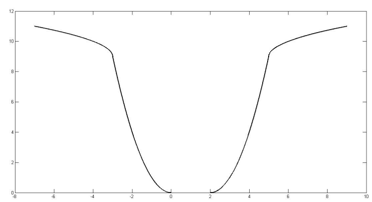

FIGURA 2.3: CUASI-CONVEXA ESTRICTA

DEFINICIÓN 2.11. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función propia. Entonces, f es llamada fuertemente Cuasi-convexa si para todo $x, y \in\ domf$ con $x \neq y$, entonces:*

$$f(\beta x + (1-\beta)y) < max\{f(x), f(y)\}\ \forall \beta \in (0,1).$$

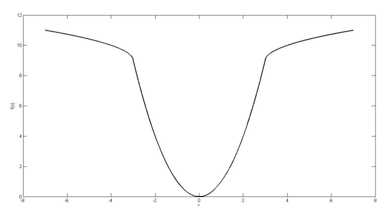

FIGURA 2.4: FUERTEMENTE CUASI-CONVEXA

PROPOSICIÓN 2.3. *La función propia $f : \mathbb{R}^n \to \mathbb{R}$ es cuasi-convexa si y solo si el conjunto de niveles $L_f(\alpha) := \{x \in domf\ :\ f(x) \leqslant \alpha\}$ es convexo $\forall \alpha \in R$.*

Demostración. Supongamos que f es cuasi-convexa, y sea $x_1, x_2 \in L_f$ así existen, α_1 y α_2 tales que $f(x_1) \leqslant \alpha_1$ y $f(x_2) \leqslant \alpha_2$ y sea $\alpha = \text{máx}\{\alpha_1, \alpha_2\}$. Sea $\lambda \in (0,1)$, y sea $x = \lambda x_1 + (1-\lambda)x_2$. Luego, por la cuasi-convexidad de f, $f(x) \leqslant \text{máx}\{f(x_1), f(x_2)\} \leqslant \alpha$. Así, $x \in L_f$ y asi L_f es convexo. Recíprocamente, supongamos que L_f es convexo para cada número real α. Sean $x_1, x_2 \in\ dom(f)$ y $\lambda \in (0,1)$. Es claro que x_1 y x_2 están en $L_f(\alpha)$ para $\alpha = \text{máx}\{f(x_1), f(x_2)\}$. Tomemos $x = \lambda x_1 + (1-\lambda)x_2$ y por la convexidad de $L_f(\alpha)$ tenemos $x \in L_f(\alpha)$, así, $f(x) \leqslant \ \alpha = \text{máx}\{f(x_1), f(x_2)\}$. Por tanto f es cuasi-convexa. ■

LEMA 2.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ una función propia inferiormente semicontinua y $\widehat{\partial} f(x)$ los subdiferenciales de f en x, entonces:*

(a) *$\widehat{\partial} f(x)$ es un conjunto cerrado y convexo.*

(b) *Si f es diferenciable en x ó en una vecindad de x, entonces, $\widehat{\partial} f(x) = \{\nabla f(x)\}$, donde $\nabla f(x)$ es el gradiente de f.*

(c) *Si $g = f + h$ con f finito en x y h diferenciable en una vecindad de x, entonces, $\widehat{\partial} g(x) = \widehat{\partial} f(x) + \nabla h(x)$.*

(d) *Si f tiene un mínimo lacal en $\bar{x}$, entonces $0 \in \widehat{\partial} f(x)$*

Demostración.

Para detalles de la demostración, ver capítulo 8 de [5].

■

PROPOSICIÓN 2.4. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función propia estrictamente cuasi-convexa. Considere el problema de minimizar $f(x)$ sujeto a $x \in C$, donde C es un conjunto convexo no vacío de $\mathbb{R}^n$. Si $\bar{x}$ es una solución óptima local, entonces $\bar{x}$ es una solución óptima global.*

Demostración. Supongamos por absurdo que existe un $\widehat{x} \in C$ tal que $f(\widehat{x}) < f(\bar{x})$. Por otro lado por la convexidad del conjunto C, $\lambda\widehat{x} + (1 - \lambda)\bar{x} \ \in \ C$ para cada $\lambda \in (0, 1)$. Luego como $\bar{x}$ es un mínimo local por hipótesis tenemos que $f(\bar{x}) \leqslant f(\lambda\widehat{x}+(1-\lambda)\bar{x})$ para todo $\lambda \in (0, \delta)$ para algún $\delta \in (0, 1)$; por otra parte, como f es estrictamente cuasi-convexa y $f(\widehat{x}) < f(\bar{x})$, se tiene que $f(\lambda\widehat{x}+(1-\lambda)\bar{x}) < f(\bar{x})$, para cada $\lambda \in (0, 1)$ pues, $\text{máx}\{f(\widehat{x}), f(\bar{x})\} = f(\bar{x})$ y este último contradice la optimalidad de $\bar{x}$.

LEMA 2.2. *Sea C un conjunto convexo no vacío de $\mathbb{R}^n$ y $f : C \to \mathbb{R}$ estrictamente cuasi-convexa e inferiormente semicontinua. Entonces f es cuasi-convexa.*

Demostración. Sean x_1 y $x_2 \in C$. Si $f(x_1) \neq f(x_2)$, entonces como f es estrictamente cuasi-convexa se tiene $f(\lambda x_1 + (1-\lambda)x_2) < \max\{f(x_1), f(x_2)\}$ para cada $\lambda \in (0,1)$ y asi estaria probado. Ahora, supongamos que $f(x_1) = f(x_2)$. Probemos que f es cuasi convexa, para ello verifiquemos que $f(\lambda x_1 + (1-\lambda)x_2) \leqslant f(x_1)$ para cada $\lambda \in (0,1)$. Supongamos por absurdo que $f(\mu x_1 + (1-\mu)x_2) > f(x_1)$ para algún $0 < \mu < 1$ y sea $x = \mu x_1 + (1-\mu)x_2$. Ahora, por hipótesis f es inferiormente semicontinua, así existe $\alpha \in (0,1)$ tal que:

$$f(x) > f(\lambda x_1 + (1-\lambda)x) > f(x_1) = f(x_2). \tag{2.31}$$

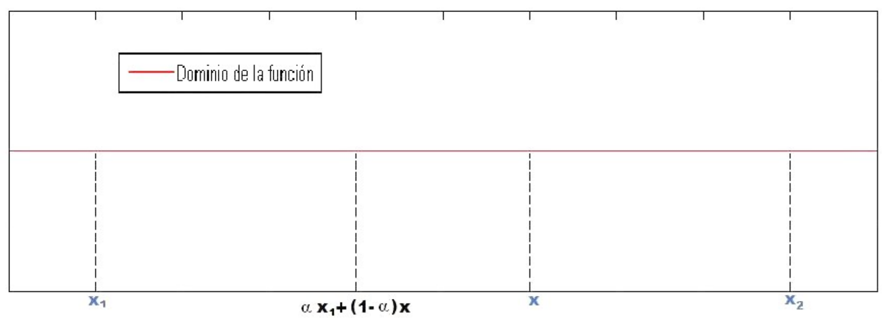

Figura 2.5: Puntos del dominio de f.

Por otra parte, observemos que x también puede ser representado como una combinación convexa entre x_2 y $\lambda x_1 + (1-\lambda)x$ ver figura(2.5), esto es, existe un $\beta \in (0,1)$ tal que:

$$x = \beta x_2 + (1-\beta)[\lambda x_1 + (1-\lambda)x].$$

Ahora, por la cuasi convexidad estricta de f, aplicando la definición (2.10) sobre la última igualdad se tiene:

$$\begin{aligned} f(x) = f(\beta x_2 + (1-\beta)[\lambda x_1 + (1-\lambda)x]) &< \max\{f(x_2), f(\lambda x_1 + (1-\lambda)x)\} \\ &= f(\lambda x_1 + (1-\lambda)x) \end{aligned}$$

con lo cual se obtiene $f(x) < f(\alpha x_1 + (1-\alpha)x)$ lo cual contradice (2.31).

■

§2.4. $\varphi-$divergencia

DEFINICIÓN 2.12 ($\varphi-$divergencia). *Sea $\varphi : \mathbb{R} \to (-\infty, +\infty]$ una función propia, convexa y cerrada. Con dom$\varphi \neq \emptyset$ y dom$\varphi \subset (0, +\infty]$, diremos que:*

(i) φ *es dos veces continuamente diferenciable en int(domφ) $= (0, +\infty)$.*

(ii) φ *es estrictamente convexa en este dominio.*

(iii) $\lim\limits_{t\to 0^+} \varphi'(t) = -\infty$.

(iv) $\varphi(1)=\varphi'(1) = 0$ *y* $\varphi''(1) > 0$.

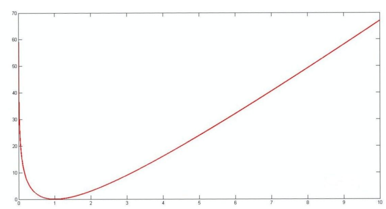

FIGURA 2.6: FUNCIÓN φ DIVERGENCIA

DEFINICIÓN 2.13. *En lo que sigue, se denotará por Φ la clase de funciones que satisfaga $(i) - (iv)$. Definamos ahora dos subclases de Φ.*

$$\Phi_1 = \{\varphi \in \Phi : \varphi''(1)(1 - \frac{1}{t}) \leqslant \varphi'(t) \leqslant \varphi''(1)\ln(t), \quad \forall t > 0\} \tag{2.32}$$

y

$$\Phi_2 = \{\varphi \in \Phi : \varphi''(1)(1 - \frac{1}{t}) \leqslant \varphi'(t) \leqslant \varphi''(1)(t-1), \quad \forall t > 0\}. \tag{2.33}$$

OBSERVACIÓN 2.1. *Como sabemos que $ln(t) \leqslant t - 1$ para todo $t > 0$ y por definición (2.12) $\varphi''(1) > 0$, se tiene que $\Phi_1 \subset \Phi_2 \subset \Phi$.*

EJEMPLOS 2.1. *Consideremos las siguientes funciones:*

$$\varphi_1(t) = t\ln(t) - t + 1, \quad dom\varphi = [0, +\infty).$$

$$\varphi_2(t) = -\ln(t) + t - 1, \quad dom\varphi = (0, +\infty].$$

Veamos que cada φ_i, con $i = \{1, 2\}$ estan en Φ_1, pues $\varphi_1'(t) = ln(t)$ y $\varphi_1''(t) = \dfrac{1}{t}$ entonces $\varphi_1''(1) = 1$, por tanto se tiene que:

$$(1 - \frac{1}{t}) \leqslant ln(t) = ln(t), \quad \forall t > 0,$$

asi se tiene que $\varphi_1 \in \Phi_1$. De manera análoga podemos ver que $\varphi_2'(t) = -\dfrac{1}{t} + 1 = 1 - \dfrac{1}{t}$ lo cual sabiendo que $\varphi_2''(1) = 1$ obtenemos:

$$1 - \frac{1}{t} = 1 - \frac{1}{t} \leqslant ln(t).$$

De esta manera $\varphi_2 \in \Phi_1$. Más aún por la observación (2.1) se tiene que $\varphi_1, \varphi_2 \in \Phi_2$.

La función φ_1 juega un papel importante en el análisis de convergencia. Esta función φ_1 es llamada tambien $Kullback - Leibler$ o función entrópica y define la siguiente casi distancia

$$H(x, y) = d_\varphi(x, y) = \sum_{j=1}^{n} y_j \varphi\left(\frac{x_i}{y_i}\right) = \sum_{j=1}^{n} x_j ln(\frac{x_j}{y_j}) + y_j - x_j, \tag{2.34}$$

donde el dominio puede ser extendido continuamente a $\mathbb{R}^n_+ \times \mathbb{R}^n_{++}$, usando como convención que $0 \ln(0) = 0$.

LEMA 2.3. *Sea $H(\cdot, \cdot)$ definido como en (2.34). Entonces se tienen los siguientes resultados.*

(a) *El conjunto de niveles de $H(x, \cdot)$ es acotado para todo $x \in \mathbb{R}^n_+$.*

(b) *Si $\{y^k\} \subset \mathbb{R}^n_{++}$ converge a $y \in \mathbb{R}^n_+$, entonces $\lim\limits_{k \to +\infty} H(y, y^k) = 0$.*

(c) *Si $\{z^k\} \subset \mathbb{R}^n_+$,$\{y^k\} \subset \mathbb{R}^n_{++}$ son sucesiones tales que $\{z^k\}$ está acotada, $\lim\limits_{k \to +\infty} y^k = y$, y además $\lim\limits_{k \to +\infty} H(z^k, y^k) = 0$, entonces el $\lim\limits_{k \to +\infty} z^k = y$.*

Demostración. **(a).** Definamos los conjuntos de nivel por:

$$\Gamma(\delta) = \{y \,:\, H(x, y) \leqslant \delta\}, \tag{2.35}$$

para todo $\delta \in \mathbb{R}_+$. Pero, probar que (2.35) es acotado es probar que:

$$\Gamma^\star(\delta) = \{y_i \,:\, H(x_i, y_i) \leqslant \delta\}, \tag{2.36}$$

para $i = 1, \ldots, n$ y $\delta \in \mathbb{R}_+$, es acotado ya que $H(x, y) \geqslant 0$, $\forall\ (x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_{++}$. Así, sean $1 \leqslant i \leqslant n$ y además x y δ fijos en $\mathbb{R}^n_+$ y $\mathbb{R}_+$ respectivamente. Y supongamos por absurdo que $\Gamma^\star(\delta)$ no es acotado, asi por definición tenemos que:

$\forall M > 0$, $\exists\, y'_i \in \Gamma^\star$ tal que $y'_i > M$, en otras palabras, $y'_i \to \infty$. Por otro lado, como $y'_i \in \Gamma^\star$ entonces se tiene que $H(x_i, y'_i) \leqslant \delta$, pero por definición de H y φ_1 se tiene:

$$\delta \geqslant H(x_i, y'_i) = y'_i \varphi \left(\frac{x_i}{y'_i} \right) \xrightarrow{y'_i \to \infty} +\infty.$$

Lo cual es una contradicción. Por tanto, $\Gamma^\star(\delta)$ es acotado para todo $i = 1, \ldots, n$ y asi, $\Gamma(\delta)$ es acotado.

(b). Usando la hipótesis y la definición de H se tiene:

$$\begin{aligned}
\lim_{k\to+\infty} H(y, y^k) &= \lim_{k\to+\infty} \sum_{j=1}^{n} (y_j ln \left(\frac{y_j}{y_j^k} \right) + y_j^k - y_j) \\
&= \sum_{j=1}^{n} \lim_{k\to+\infty} (y_j ln \left(\frac{y_j}{y_j^k} \right) + y_j^k - y_j) \\
&= \sum_{j=1}^{n} (y_j ln \left(\frac{y_j}{y_j} \right) + y_j - y_j) \qquad\qquad [ya \;\; que \;\; y^k \to y] \\
&= \sum_{j=1}^{n} (y_j ln(1) + y_j - y_j) \\
&= \sum_{j=1}^{n} (y_j 0 + 0) \\
&= \sum_{j=1}^{n} (0) \\
&= 0.
\end{aligned}$$

(c). Por hipótesis tenemos:

$$
\begin{aligned}
\lim_{k\to\infty} H(z^k, y^k) = 0 \Longrightarrow & \lim_{k\to\infty} \sum_{i=1}^{n} y_i^k \varphi_1\left(\frac{z_i^k}{y_i^k}\right) = 0 \\
\Longrightarrow & \sum_{i=1}^{n} \lim_{k\to\infty} y_i^k \varphi_1\left(\frac{z_i^k}{y_i^k}\right) = 0 \\
\Longrightarrow & \lim_{k\to\infty} y_i^k \varphi_1\left(\frac{z_i^k}{y_i^k}\right) = 0 \\
\Longrightarrow & \lim_{k\to\infty} \varphi_1\left(\frac{z_i^k}{y_i^k}\right) = 0. \qquad\qquad y_i^k \neq 0 \qquad\qquad (2.37) \\
\Longrightarrow & \varphi_1(\lim_{k\to\infty} \frac{z_i^k}{y_i^k}) = 0 \\
\Longrightarrow & \lim_{k\to\infty} \frac{z_i^k}{y_i^k} = 1 \\
\Longrightarrow & \lim_{k\to\infty} z_i^k = \lim_{k\to\infty} y_i^k \\
\Longrightarrow & \lim_{k\to\infty} z_i^k = y.
\end{aligned}
$$

La desigualdad anterior también es válida, pues en el caso en que $y = 0$, retomando la parte izquierda de la desigualdad (2.37) se tiene:

$$\lim_{k\to\infty} y_i^k \varphi_1\left(\frac{z_i^k}{y_i^k}\right),$$

tiene forma indeterminada $0.\infty$ debido a que φ es continua, cambiando la expresión anterior por:

$$\lim_{k\to\infty} \frac{\varphi_1\left(\frac{z_i^k}{y_i^k}\right)}{\frac{1}{y_i^k}},$$

ahora se tiene la forma indeterminada ∞/∞, aplicando la regla de L'Hôpital a la última expresión:

$$\lim_{k\to\infty} \frac{\varphi_1'\left(\frac{z_i^k}{y_i^k}\right)\left(-\frac{z_i^k}{(y_i^k)^2}\right)}{-\frac{1}{(y_i^k)^2}} = \lim_{k\to\infty} z_i^k \varphi_1'\left(\frac{z_i^k}{y_i^k}\right).$$

Por otra parte, por hipótesis se tiene que $\{z^k\}_{k\in\mathbb{N}}$ es acotada, entonces solo resta calcular:

$$\lim_{k\to\infty} \varphi_1'\left(\frac{z_i^k}{y_i^k}\right).$$

Luego, usando la continuidad de φ_1', la expresión $\dfrac{z_i^k}{y_i^k} \stackrel{k\to\infty}{\longrightarrow} +\infty$ y asi se tiene que

$$\lim_{k\to\infty} \varphi_1'\left(\frac{z_i^k}{y_i^k}\right) = +\infty,$$

Lo cual es una contradicción, ya que por hipótesis $\lim_{k\to\infty} y_i^k \varphi_1\left(\dfrac{z_i^k}{y_i^k}\right) = 0.$

DEFINICIÓN 2.14 (Kernel ϕ). *Sean $\varphi \in \Phi$ y los parámetros $\mu > 0$, $\nu \geqslant 0$. Se define como kernel, a la función propia convexa cerrada $\phi : \mathbb{R} \to (-\infty, +\infty]$ dada por:*

$$\phi(t) = \mu\varphi(t) + \frac{\nu}{2}(t-1)^2 \tag{2.38}$$

OBSERVACIONES 2.2.

1) $\phi \in \Phi$. Veamos que se satisfacen las 4 condiciones de la definición (2.12).

(i) Es satisfecha, pues por hipótesis $\varphi \in \Phi$ y el término cuadrático $(t-1)^2$ es un polinomio y así se tiene que ϕ es dos veces continuamente diferenciable.

(ii) Se cumple pues φ es estrictamente convexa y el término cuadrático es convexo, además como los parámetos ν y μ son no negativos se tiene que esta función también es estrictamente convexa.

(iii) Es válida para ϕ pues la derivada del término cuadrático $(t-1)^2$ tiende a -2 por la continuidad y como $\varphi'(t) \to -\infty$ cuando $t \to 0^+$ entonces se tiene que

$$\lim_{t\to 0^+} \phi'(t) = \lim_{t\to 0^+} [\mu\varphi'(t) + \nu(t-1)] = -\infty.$$

Además $\phi(1) = \mu\varphi(1) + \dfrac{\nu}{2}(1-1)^2 = \mu\varphi(1) = 0 = \mu\varphi'(1) + \nu(1-1) = \mu\varphi'(1)$ y $\phi''(t) = \mu\varphi''(t) + \nu > 0$ pues $\nu > 0$. Con lo cual se satisface (iv) y asi $\phi \in \Phi$.

2) ϕ es fuertemente convexa, si $v > 0$. Sabemos las siguientes caracterizaciones para funciones convexas en $\mathbb{R}$. Si f es dos veces continuamente diferenciable y el dominio es la recta real, entonces podemos caracterizar de la siguiente manera:

1. *f convexa si y solo si para todo $f''(x) \geqslant 0$.*

2. f *estrictamente convexa si* $f''(x) > 0$ *para todo* x.

3. f *fuertemente convexa si y solo si* $f''(x) \geqslant m > 0$ *para todo* x.

Asi, como φ *es estrictamente convexa entonces* $\varphi''(t) \geqslant 0$ *para todo* t *y asi* $\phi''(t) = \mu\varphi''(t) + \nu \geqslant \nu > 0$ *para todo* t *con lo cual se tiene que* φ *es estrictamente convexa por la caracterización (3) enumerada anteriormente.*

3) La función objetivo del problema inicial

$$f(x) + \lambda \sum_{i=1}^{n} (x_i^k)^2 \phi\left(\frac{x_i}{x_i^k}\right) \tag{2.39}$$

Puede ser estrictamente convexa en $\mathbb{R}^n_{++}$ *para algún* λ *suficientemente grande, a pesar de que* f *sea cuasi-convexa.*

Definición 2.15. *Para la definición dada en (2.14) se define la casi distancia:*

$$d_\phi(x, y) = \sum_{i=1}^{n} (y_i)^2 \phi\left(\frac{x_i}{y_i}\right)$$

Lema 2.4. *Sea* $\varphi \in \phi$*, los parámetros* $\mu > 0$*,* $\nu \geqslant 0$*, y* ϕ *el kernel definido como en (2.14) y* $d_\phi(\cdot, \cdot)$ *la función inducida por* ϕ *dada en (2.15). entonces:*

(a) d_ϕ *es una función homogénea de segundo orden.*

(b) *Para* $y \in \mathbb{R}^n_{++}$ *fijo, la función* $d_\phi(\cdot, y)$ *es estrictamente convexa sobre* $\mathbb{R}^n_{++}$*. Si adicionamos,* $\nu > 0$*, entonces* $d_\phi(\cdot, y)$ *es fuertemente convexa en* $\mathbb{R}^n_{++}$*.*

(c) *Para cada* $(x, y) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$*,* $d_\phi(x, y) \geqslant 0$ *y* $d_\phi(x, y) = 0$ *si y solo si* $x = y$*.*

(d) *Para cada* $z \in \mathbb{R}^n_{++}$ *fijo, el conjunto de niveles* $L(z, \gamma) := \{x \in \mathbb{R}^n_{++} : d_\phi(x, z) \leqslant \gamma\}$ *son acotados para todo* $\gamma \geqslant 0$*.*

(e) *Si* $\varphi \in \Phi_1$ *ó* Φ_2*, y* $\{y^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n_{++}$ *converge a* $\bar{y} \in \mathbb{R}^n_{+}$*, entonces para algún* $x \in \mathbb{R}^n_{++}$ *fijo, la sucesión* $\{d_\phi(x, y^k)\}_{k\in\mathbb{N}}$ *está acotada.*

Demostración. **(a)** Por definición de d_ϕ se tiene:

$$\begin{aligned} d_\phi(\alpha x, \alpha y) &= \sum_{i=1}^{n} (\alpha y_i)^2 \phi(\frac{\alpha x_i}{\alpha y_i}) \\ &= \sum_{i=1}^{n} \alpha^2 y_i^2 \phi(\frac{x_i}{y_i}) \\ &= \alpha^2 \sum_{i=1}^{n} y_i^2 \phi(\frac{x_i}{y_i}) \\ &= \alpha^2 d_\phi(x, y). \end{aligned}$$

(b) Sea y fijo en $\mathbb{R}^n_{++}$, además sabemos que ϕ es estrictamente convexa por la observación (2.2)(ii). Luego, usando la definición de d_ϕ se tiene, para x arbitrario y $\alpha \in (0, 1)$:

$$\begin{aligned} d_\phi(\alpha x + (1-\alpha)\bar{x}, y) &= \sum_{i=1}^{n} y_i^2 \phi\left(\frac{\alpha x_i + (1-\alpha)\bar{x}_i}{y_i}\right) \\ &= \sum_{i=1}^{n} y_i^2 \phi\left(\alpha\frac{x_i}{y_i} + (1-\alpha)\frac{\bar{x}_i}{y_i}\right) \\ &< \sum_{i=1}^{n} \left[y_i^2 \alpha \phi\left(\frac{x_i}{y_i}\right) + (1-\alpha) y_i^2 \phi\left(\frac{\bar{x}_i}{y_i}\right)\right] \\ &= \alpha \sum_{i=1}^{n} y_i^2 \phi\left(\frac{x_i}{y_i}\right) + (1-\alpha) \sum_{i=1}^{n} y_i^2 \phi\left(\frac{\bar{x}_i}{y_i}\right) \\ &= \alpha d_\phi(x, y) + (1-\alpha) d_\phi(\bar{x}, y). \end{aligned}$$

Así, obtenemos que $d_\phi(\cdot, y)$ es estrictamente convexa.

(c) Por la observación (2.2) (i) se tiene que $\phi \in \Phi$ y por tanto es estrictamente convexa. Además como la derivada de $\phi'(1) = \phi(1) = 0$ se tiene un mínimo global, es decir:

$$\phi(t) \geqslant \phi(1) = 0, \quad \forall t.$$

Luego, se tiene que $\phi(t) = 0$ si y solo si $t = 1$.

$\therefore$ $d_\phi(x, y) \geqslant 0, \forall (x, y)$ en $\mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ y además al ser una suma de números no negativos, se tiene que $d_\phi(x, y) = 0$ si y solo si $x = y$.

(d). Para probar este resultado, solo es necesario considerar el caso unidimensional, es decir, probar que $h_\xi(t) = \xi^2\phi(\frac{t}{\xi}), \forall \xi > 0$ y ver si el conjunto de niveles está acotado, pero esto a su vez, es ver que ϕ tiene un conjunto de niveles acotado, pues ξ es fijo. Pero por definición se tiene que $\{t : \phi(t) \leqslant 0\} = \{1\}$. Utilizando el corolario 8.7.1 de [4]. se tiene que el conjunto de niveles de ϕ es acotado y por tanto se tiene lo deseado.

(e) Por las definiciones de ϕ y d_ϕ se tiene que:

$$\begin{aligned} d_\phi(x, y^k) &= \sum_{i=1}^{n}\left[\mu(y_i^k)^2\varphi\left(\frac{x_i}{y_i^k}\right) + \frac{\nu}{2}(y_i^k)^2\left(\frac{x_i}{y_i^k} - 1\right)^2\right] \\ &= \sum_{i=1}^{n}\left[\mu(y_i^k)^2\varphi\left(\frac{x_i}{y_i^k}\right) + \frac{\nu}{2}(x_i - y_i^k)^2\right]. \end{aligned}$$

Luego si φ está acotada para todo $t > 0$ no tendriamos nada que probar, pues por la continuidad del término cuadrático se tendría lo deseado. Veamos ahora los 2 casos siguientes:

Caso i. si $\bar{y}_i > 0$ para todo $i = \{1, \cdots, n\}$. Entonces como por hipótesis se tiene que $\{y_i^k\}_{k\in\mathbb{N}} \to y_i^k$, para cada i. entonces por la continuidad de φ se tiene la prueba directamente.

Caso ii. Si existiera un índice $i_0 \in \{1, \cdots, n\}$ tal que $y_{i_0} = 0$ será suficiente probar que la sucesión $\{(y_{i_0}^k)^2\varphi(\frac{x_i}{y_{i_0}^k})\}$ está acotada.

Para cada $k \in \mathbb{N}$, al usar la convexidad, diferenciabilidad y el hecho de que $\varphi(1) = 0$. se tiene que:

$$\varphi(1) \geqslant \varphi\left(\frac{x_i}{y_{i_0}^k}\right) + \varphi'\left(\frac{x_i}{y_{i_0}^k}\right)\left(1 - \frac{x_i}{y_{i_0}^k}\right) \Leftrightarrow 0 \geqslant \varphi\left(\frac{x_i}{y_{i_0}^k}\right) + \varphi'\left(\frac{x_i}{y_{i_0}^k}\right)\left(1 - \frac{x_i}{y_{i_0}^k}\right). \tag{2.40}$$

Así, multiplicando (2.40) por $(y_{i_0}^k)^2$ y reordenando un poco los términos, se tiene:

$$\begin{aligned}(y_{i_0}^k)^2\varphi\left(\frac{x_i}{y_{i_0}^k}\right) &\leqslant (y_{i_0}^k)^2\varphi'\left(\frac{x_i}{y_{i_0}^k}\right)\left(\frac{x_i}{y_{i_0}^k}-1\right)\\ &= (y_{i_0}^k)^2\varphi'\left(\frac{x_i}{y_{i_0}^k}\right)(x_i-y_{i_0}^k)\\ &\leqslant \left|(y_{i_0}^k)^2\varphi'\left(\frac{x_i}{y_{i_0}^k}\right)(x_i-y_{i_0}^k)\right| \qquad (2.41)\end{aligned}$$

Por otro lado, si $\varphi \in \Phi_2$ entonces sigue de (2.33) que:

$$\varphi''(1)(1-\frac{y_{i_0}^k}{x_i}) \leqslant \varphi'(\frac{x_i}{y_{i_0}^k}) \leqslant \varphi''(1)(\frac{x_i}{y_{i_0}^k}-1) \Leftrightarrow \varphi''(1)y_{i_0}^k(1-\frac{y_{i_0}^k}{x_i}) \leqslant y_{i_0}^k\varphi'(\frac{x_i}{y_{i_0}^k}) \leqslant \varphi''(1)(x_i-y_{i_0}^k)$$

Ahora, multiplicando por $(x_i - y_{i_0}^k)$ tenemos:

<u>Caso 1:</u> $(x_i - y_{i_0}^k) > 0$

$$\varphi''(1)\frac{y_{i_0}^k}{x_i}(x_i-y_{i_0}^k)^2 \leqslant y_{i_0}^k(x_i-y_{i_0}^k)\varphi'(\frac{x_i}{y_{i_0}^k}) \leqslant \varphi''(1)(x_i-y_{i_0}^k)^2$$

<u>Caso 2:</u> $(x_i - y_{i_0}^k) < 0$

$$\varphi''(1)(x_i-y_{i_0}^k)^2 \leqslant y_{i_0}^k(x_i-y_{i_0}^k)\varphi'(\frac{x_i}{y_{i_0}^k}) \leqslant \varphi''(1)\frac{y_{i_0}^k}{x_i}(x_i-y_{i_0}^k)^2$$

$\therefore$ Al considerar ambos casos y (2.41) podemos concluir que:

$$(y_{i_0}^k)^2\varphi(\frac{x_i}{y_{i_0}^k}) \leqslant \text{máx}\left\{\varphi''(1)(x_i-y_{i_0}^k)^2, \varphi''(1)\frac{y_{i_0}^k}{x_i}(x_i-y_{i_0}^k)^2\right\}$$

Así, junto con las hipótesis dadas anteriormente se obtiene que $\{(y_{i_0}^k)^2\varphi(\frac{x_i}{y_{i_0}^k})\}$, está acotado para todo $\varphi \in \Phi_2$ y en consecuencia la sucesión $\{d_\phi(x, y^k)\}_{k\in\mathbb{N}}$ está acotada. Por otra parte, como sabemos que $\Phi_1 \subset \Phi_2$ tenemos así que la sucesión $\{d_\phi(x, y^k)\}_{k\in\mathbb{N}}$ es acotada para $\varphi \in \Phi_1$. ∎

DEFINICIÓN 2.16. *En analogía con la distancia euclideana, se puede definir la proyección de un punto y, denotado por $\widehat{x}(y)$, al conjunto convexo cerrado $S \subset \mathbb{R}^n$ con respecto a d_ϕ, como la solución del siguiente problema*

$$\inf\{d_\phi(x, y) : x \in S\}.$$

La existencia de $\widehat{x}(y)$ está garantizada por el lema 2.4 (d).

Definición 2.17. *Un vector x^* se dice que es normal a un conjunto convexo C en un punto $a \in C$ si:*

$$\langle x - a, x^* \rangle \leqslant 0, \quad \forall x \in C.$$

Lema 2.5. *Sea S un subconjunto convexo cerrado de $\mathbb{R}^n$, $y \in S^c$. entonces $\widehat{x}(y)$ es la proyección de y en S con respecto a d_ϕ si y solo si*

$$\langle x - \widehat{x}(y), -\nabla d_\phi(\widehat{x}(y), y) \rangle \leqslant 0, \quad \forall x \in S. \tag{2.42}$$

Demostración. El problema descrito en la definición (2.16) se puede reescribir como $\inf\{d_\phi(x, y) + \delta(x \mid S) : x \in \mathbb{R}^n\}$, donde $\delta(x \mid S)$ denota la función indicadora de el conjunto S. Ahora, por teorema 27.4 de [4] y como por hipótesis $\widehat{x}(y)$ es solución de ese problema se tiene:

$$\begin{aligned} 0 \in \partial\{d_\phi(x, y) + \delta(x \mid S)\} &\Longleftrightarrow 0 \in \{\nabla d_\phi(x, y) + \partial\delta(x \mid S)\} \\ &\Longleftrightarrow -\nabla d_\phi(x, y) \in \partial\delta(x \mid S) \end{aligned} \tag{2.43}$$

Ahora, por definición de subgradiente, aplicandola a la función indicadora y tomando en cuenta (2.43) se obtine:

$$\langle x - \widehat{x}(y), -\nabla d_\phi(\widehat{x}(y), y) \rangle \leqslant 0, \quad \forall x \in S.$$

Lo cual, por la definición (2.17) obtenemos lo deseado. ■

Lema 2.6. *Dado $\varphi \in \Phi$ y los parámetros $\mu > 0$, $\nu \geqslant 0$, sea ϕ el kernel definido como en (2.14). Entonces, para todo $a, b \in \mathbb{R}^n_{++}$ y $c \in \mathbb{R}^n_+$, se tienen los siguientes resultados:*

(a) *Si $\nu > 0$ y $\varphi \in \Phi_1$, entonces*

$$\langle c - b, \nabla_1 d_\phi(b, a) \rangle \leqslant \mu\varphi''(1) \max_{1 \leqslant j \leqslant n} \{a_j\}[H(c, a) - H(c, b)].$$

.

(b) *Si $\nu \geqslant \mu\varphi''(1) > 0$ y $\varphi \in \Phi_2$, entonces*

$$\langle c - b, \nabla_1 d_\phi(b, a) \rangle \leqslant \theta(\|c - a\|^2 - \|c - b\|^2), \quad \text{donde } \theta = \frac{\nu + \mu\varphi''(1)}{2}$$

Demostración.

(a). Por definición, si $\varphi \in \Phi_1$ entonces:

$$\varphi'(t) \leqslant \varphi''(1)\ln(t), \quad \forall t > 0.$$

Tomando ahora, $t = \dfrac{b_j}{a_j}$ y multiplicando por $(c_j a_j)$, la desigualdad anterior tenemos:

$$c_j a_j \varphi'(\frac{b_j}{a_j}) \leqslant c_j a_j \varphi''(1)\ln(\frac{b_j}{a_j}), \quad j = 1, 2, \ldots, n. \tag{2.44}$$

Ahora, tomando la otra parte de la desigualdad de la definición (2.32) tenemos:

$$-\varphi'(t) \leqslant -\varphi''(1)(1 - \frac{1}{t}),$$

y de igual manera, sustituyendo $t = \dfrac{b_j}{a_j}$ y multiplicando por a_j se tiene:

$$\begin{aligned} -a_j\varphi'(\frac{b_j}{a_j}) &\leqslant -a_j\varphi''(1)(1 - \frac{a_j}{b_j}) \\ &= -\frac{a_j}{b_j}\varphi''(1)(b_j - a_j), \end{aligned}$$

con lo cual se obtiene:

$$-a_j b_j \varphi'(\frac{b_j}{a_j}) \leqslant -a_j\varphi''(1)(b_j - a_j) \quad j = 1, 2, \ldots, n. \tag{2.45}$$

Definamos:

$$\psi'(a, b) = \left(a_1\varphi'(\frac{b_1}{a_1}), \ldots, a_n\varphi'(\frac{b_n}{a_n})\right)^T, \quad \forall a, b \in \mathbb{R}^n_{++}.$$

Probemos ahora por definición que $\mu\psi'(a, b) = \nabla_1 d_\phi(b, a)$.

$$\begin{aligned} \nabla_1 d_\phi(x, y) &= \nabla_1 \left[\sum_{i=1}^{n} (y_i)^2 \phi(x_i/y_i)\right] \\ &= \nabla_1 \left[\sum_{i=1}^{n} (y_i)^2 \mu\ \varphi\left(\frac{x_i}{y_i}\right)\right] \\ &= \left(\mu y_1\ \varphi'\left(\frac{x_1}{y_1}\right), \mu y_2\ \varphi'\left(\frac{x_2}{y_2}\right), \ldots, \mu y_n\ \varphi'\left(\frac{x_n}{y_n}\right)\right)^T. \end{aligned}$$

Así evaluando en (b,a) se tiene:

$$\begin{aligned}\nabla_1 d_\phi(b,a) &= \left(\mu a_1 \ \varphi'\left(\frac{b_1}{a_1}\right), \mu a_2 \ \varphi'\left(\frac{b_2}{a_2}\right), \dots, \mu a_n \ \varphi'\left(\frac{b_n}{a_n}\right)\right)^T \\ &= \mu \ \psi'(a,b). \end{aligned} \tag{2.46}$$

Ahora usando las desigualdades (2.44), (2.45), sobre cada $i = 1, \dots, n$ se tiene:

$$\begin{aligned}
\langle c-b, \psi'(a,b)\rangle &= \sum_{i=1}^{n}(c_i - b_i)a_i\varphi'\left(\frac{b_i}{a_i}\right) \\
&= \sum_{i=1}^{n}\left(c_i a_i\varphi'\left(\frac{b_i}{a_i}\right) - b_i a_i \varphi'\left(\frac{b_i}{a_i}\right)\right) \\
&\leqslant \sum_{i=1}^{n}\left(c_i a_i\varphi''(1)\ln\left(\frac{b_i}{a_i}\right) - b_i a_i \varphi'\left(\frac{b_i}{a_i}\right)\right) \\
&\leqslant \sum_{i=1}^{n}\left(c_i a_i\varphi''(1)\ln\left(\frac{b_i}{a_i}\right) + a_i\varphi''(1)(a_i - b_i)\right) \\
&= \varphi''(1)\sum_{i=1}^{n} a_i\left(c_i\ln\left(\frac{b_i}{a_i}\right) + (a_i - b_i)\right) \\
&\leqslant \varphi''(1)\max_{1\leqslant i\leqslant n}\{a_i\}\left[\sum_{i=1}^{n} c_i\ln\left(\frac{b_i}{a_i}\right) + (a_i - b_i)\right] \\
&= \varphi''(1)\max_{1\leqslant i\leqslant n}\{a_i\}\left[\sum_{i=1}^{n} c_i(\ln(b_i) - \ln(a_i) - \ln(c_i) + \ln(c_i)) + a_i - b_i + c_i - c_i\right] \\
&= \varphi''(1)\max_{1\leqslant i\leqslant n}\{a_i\}\left[\sum_{i=1}^{n} c_i(\ln(c_i) - \ln(a_i) - [\ln(c_i) - \ln(b_i)]) + a_i - b_i + c_i - c_i\right] \\
&= \varphi''(1)\max_{1\leqslant i\leqslant n}\{a_i\}\left[\sum_{i=1}^{n} c_i\ln\left(\frac{c_i}{a_i}\right) - c_i\ln\left(\frac{c_i}{b_i}\right) + a_i - b_i + c_i - c_i\right] \\
&= \varphi''(1)\max_{1\leqslant i\leqslant n}\{a_i\}\sum_{i=1}^{n}\left[c_i\ln\left(\frac{c_i}{a_i}\right) + a_i - c_i\right] - \left[c_i\ln\left(\frac{c_i}{b_i}\right) + b_i - c_i\right] \\
&= \varphi''(1)\max_{1\leqslant i\leqslant n}\{a_i\}\left[\sum_{i=1}^{n} H(c,a) - H(c,b)\right].
\end{aligned}$$

Luego al multiplicar por μ y usando la igualdad (2.46) se tiene lo deseado.

(b) Como $\varphi \in \Phi_2$ por (2.33) se tiene:

$$\varphi'(t) \leqslant \varphi''(1)(t-1), \quad \forall t > 0.$$

Luego, sustituyendo $t = \frac{b_i}{a_i}$ y multiplicando por c_i se tiene:

$$c_i a_i \varphi'(\frac{b_i}{a_i}) \leqslant c_i a_i \varphi''(1) \left(\frac{b_i}{a_i} - 1\right) = c_i\, \varphi''(1)(b_i - a_i), \quad i = 1, \dots, n. \tag{2.47}$$

Ahora usando (2.47) y (2.45) y la definición de ψ', sobre $i = 1, \dots, n$ tenemos:

$$\begin{aligned}
\langle c - b, \psi'(a,b)\rangle &= \sum_{i=1}^{n} \left(c_i a_i \varphi'\left(\frac{b_i}{a_i}\right) - b_i a_i \varphi'\left(\frac{b_i}{a_i}\right)\right) \\
&\leqslant \sum_{i=1}^{n} c_i\, \varphi''(1)(b_i - a_i) - a_i \varphi''(1)(b_i - a_i) \\
&= \varphi''(1) \sum_{j=1}^{n} c_i\,(b_i - a_i) - a_j(b_i - a_i) \\
&= \varphi''(1)\langle c - a, b - a\rangle.
\end{aligned}$$

Como por hipótesis $\nu \neq 0$, se tiene que $\nabla_1 d_\phi(b,a) = \mu\psi'(a,b) - \nu(b-a)$, usando este resultado en la última desigualdad se tiene:

$$\langle c - b, \nabla_1 d_\phi(a,b)\rangle \leqslant \mu\varphi''(1)\langle c - a, b - a\rangle + \nu\langle c - b, b - a\rangle. \tag{2.48}$$

Haciendo uso de las siguentes igualdades:

$$\langle c - a, b - a\rangle = \frac{1}{2}(\|c-a\|^2 - \|c-b\|^2 + \|b-a\|^2)$$

y

$$\langle c - b, b - a\rangle = \frac{1}{2}(\|c-a\|^2 - \|c-b\|^2 - \|b-a\|^2).$$

Sustituyendo en (2.48) obtenemos:

$$\begin{aligned}
\langle c-b, \nabla_1 d_\phi(b,a)\rangle &\leqslant \frac{1}{2}(\|c-a\|^2 - \|c-b\|^2 + \|b-a\|^2)\mu\varphi''(1) + \\
&\quad \nu(\frac{1}{2}(\|c-a\|^2 - \|c-b\|^2 - \|b-a\|^2)) \\
&= \frac{1}{2}(\mu\varphi''(1)+\nu)[\|c-a\|^2 - \|c-b\|^2] - \frac{1}{2}(\nu - \mu\varphi''(1)\|b-a\|^2) \\
&= \theta(\|c-a\|^2 - \|c-b\|^2) - \frac{1}{2}(\nu - \mu\varphi''(1)\|b-a\|^2) \\
&\leqslant \theta(\|c-a\|^2 - \|c-b\|^2) \qquad\qquad [Pues\ \nu \geqslant \mu\varphi''(1)].
\end{aligned}$$

■

Capítulo 3

Métodos de punto proximal

En este capítulo se consideran dos clases de algoritmos de puntos proximales, basados en la función casi distancia homogénea d_ϕ de segundo orden para el problema (5) de optimización cuasi-convexa. Estos algoritmos son descritos bajo las siguientes hipótesis:

§3.1. Método Proximal Interior(IPM)

Sea ϕ como en la definición (2.14), con $\mu > 0$, $\nu = 0$ y $\varphi \in \Phi_1$, generando la sucesión $\{x^k\}_{k\in\mathbb{N}}$ dada en el esquema (6).

§3.2. Método Proximal Interior Regularizado (RIPM)

Sea ϕ definido como en (2.14), con $\mu > 0$,$\nu \geqslant \mu\varphi''(1)$ y $\varphi \in \Phi_2$, generando la sucesión $\{x^k\}_{k\in\mathbb{N}}$ dada en el esquema (6).

Con el fin de establecer la convergencia de los métodos, se tomarán las siguientes hipótesis para el problema de optimización cuasi-convexa (5):

(A1) $dom f \bigcap \mathbb{R}^n_{++} \neq \emptyset$.

(A2) El conjunto óptimo del problema (5), el cual denotamos por X^* es no vacío y acotado.

En lo siguiente, se probará la convergencia de los métodos (IPM) y (RIPM).

Veamos entonces que la sucesión generada por (6) está bien definida; esto es consecuencia del siguiente lema.

Lema 3.1. *Dados $\mu > 0$, $\nu \geqslant 0$ y $\varphi \in \Phi$, además sean ϕ y d_ϕ como en las definiciones (2.14) y (2.15), respectivamente. Entonces, bajo las hipótesis **(A1)** y **(A2)**, la sucesión generada por el esquema iterativo (6) está bien definida.*

Demostración.

Se procederá por inducción. Para $k = 0$, es válido pues por hipótesis **(A1)**, $dom f \cap \mathbb{R}^n \neq \emptyset$. De esta manera existe $x^0 > 0$.

Veamos ahora que para el término $k+1$ también se cumple.

Sea f^* un valor óptimo del problema (5), luego:

$$f^* \leqslant f(x),\ \forall x \in \mathbb{R}^n_{++} \Leftrightarrow f^* + \lambda_k d_\phi(x, x^k) \leqslant f(x) + \lambda_k d_\phi(x, x^k),\ \forall x \in \mathbb{R}^n_{++}. \quad (3.1)$$

Sea $F_k(x) = f(x) + \lambda_k d_\phi(x, x^k)$ y así:

$$f^* + \lambda_k d_\phi(x, x^k) \leqslant F_k(x).$$

Ahora, denotemos el conjunto de nivel por:

$$L_{F_k}(\gamma) = \{x \in \mathbb{R}^n_{++} : F_k(x) \leqslant \gamma\},\ \textit{para todo}\ \gamma \in \mathbb{R}. \quad (3.2)$$

Además, por (3.1) se tiene que:

$$\begin{aligned} F_k(x) \leqslant \gamma &\Leftrightarrow f(x) + \lambda_k d_\phi(x, x^k) \leqslant \gamma \\ &\Leftrightarrow f^* + \lambda_k d_\phi(x, x^k) \leqslant \gamma \\ &\Leftrightarrow d_\phi(x, x^k) \leqslant (\gamma - f^*)\lambda_k^{-1} \\ &\Leftrightarrow x \in L(x^k, (\gamma - f^*)\lambda_k^{-1}). \end{aligned}$$

$\therefore L_{F_k}(\gamma) \subset L(x^k, \lambda_k^{-1}(\gamma - f^*))$. Luego, por el lema 2.4(d), el conjunto de niveles $L(x^k, \lambda_k^{-1}(\gamma - f^*))$ es acotado, para $\lambda_k^{-1}(\gamma - f^*) \geqslant 0 \Leftrightarrow \gamma \geqslant f^*$. Por otro lado, veamos que $L_{F_k} \subset X^*$.

Sea $y \in L_{F_k(\gamma)}$, luego por definición (3.2) se tiene:

$$\begin{aligned} F_k(y) \leqslant \gamma &\Longrightarrow f(y) + \lambda_k d_\phi(y, x^k) \leqslant \gamma \\ &\Longrightarrow f(y) \leqslant f(y) + \lambda_k d_\phi(y, x^k) \leqslant \gamma \\ &\Longrightarrow f(y) \leqslant \gamma \qquad (3.3) \\ &\Longrightarrow y \in X^* \qquad . \end{aligned}$$

Luego con este resultado, considerando el caso $\gamma \leqslant f^*$, por (3.3) se tiene que $f(y) \leqslant \gamma \leqslant f^*$, lo cual implica que $y \in X^*$ pues, f^* es un valor óptimo. Así, podemos decir que el conjunto $L_{F_k}(\gamma)$ es acotado para todo $\gamma \in \mathbb{R}$. Por otra parte, $F_k(x)$ es semicontinua inferior en $dom(f)$ y por tanto el conjunto de nivel $L_{F_k}(\gamma)$ es compacto para todo $\gamma \in \mathbb{R}$.

Por último, usando la semicontinuidad inferior de $F_k(x)$ y la compacidad del conjunto de nivel, obtenemos que $F_k(x)$ tiene un mínimo global, pero este no puede ser único, debido a la no convexidad de f. En este caso x^{k+1} puede ser tomado de manera arbitraria en el conjunto de minimizadores de $F_k(x)$.

■

Veamos algunas propiedades para la sucesión $\{x^k\}_{k\in\mathbb{R}}$ bajo las hipótesis anteriormete mencionadas para utilizarlas más adelante en la convergencia de los métodos (IPM) y (RIPM).

Definición 3.1. *Definamos el siguiente conjunto el cual será de gran utilidad para la convergencia de los métodos.*

$$U = \{x \in \mathbb{R}^n_+ / f(x) \leqslant \inf_{k\in\mathbb{N}} f(x^k)\}.$$

Por la hipótesis (A1) y (A2), U es no vacío. Veamos que U es convexo, sea $x_1, x_2 \in U$ y $\lambda \in (0,1)$, Así existen $l, j \in \mathbb{N}$ tales que $f(x_1) \leqslant \inf_{l\in\mathbb{N}} f(x^l)$ y $f(x_2) \leqslant \inf_{j\in\mathbb{N}} f(x^j)$.

Sea $\inf_{k\in\mathbb{N}} f(x^k) = \max\left\{\inf_{l\in\mathbb{N}} f(x^l), \inf_{j\in\mathbb{N}} f(x^j)\right\}$. Ahora, por hipótesis f es cuasiconvexa luego:

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \max\{f(x_1), f(x_2)\} \leqslant \inf_{k\in\mathbb{N}} f(x^k)$$

.

Además U cerrado, pues si tomamos una sucesión en U, digamos $y_n \to x$ por la semicontinuidad de f obtenemos:

$$\begin{aligned} & f(x) \leqslant \liminf_{y\to x} f(y) \leqslant \liminf f(y_n) \leqslant \inf_{k\in\mathbb{N}} f(x^k) \\ \Longrightarrow & f(x) \leqslant \inf_{k\in\mathbb{N}} f(x^k) \\ \Longrightarrow & x \in U. \end{aligned} \tag{3.4}$$

Asi, U es no vacío, convexo y cerrado.

LEMA 3.2. *Sea $\{\lambda_k\}_{k\in\mathbb{N}}$ una sucesión arbitraria de números positivos y $\{x^k\}_{k\in\mathbb{N}}$ la sucesión generada por IPM. Entonces, bajo las hipótesis (A1) y (A2):*

(a) $\{f(x^k)\}_{k\in\mathbb{N}}$ es una sucesión decreciente y convergente.

(b) $\{x^k\}_{k\in\mathbb{N}}$ es Fejér convergente en el conjunto U con respecto a H definida como en (2.34).

(c) $\forall x \in U$, la sucesión $\{H(x, x^k)\}_{k\in\mathbb{N}}$ es convergente.

Demostración.

(a) Por la ecuación (6), x^{k+1} es una solución óptima global del problema:

$$\min_{x\geqslant 0}\{f(x) + \lambda_k d_\phi(x, x^k)\},$$

y en consecuencia para $x \in \mathbb{R}^n_+$, se tiene que:

$$f(x^{k+1}) + \lambda_k d_\phi(x^{k+1}, x^k) \leqslant f(x) + \lambda_k d_\phi(x, x^k). \tag{3.5}$$

Luego, sustituyendo $x = x^k$ en (3.5) tenemos:

$$\begin{aligned}
& f(x^{k+1}) + \lambda_k d_\phi(x^{k+1}, x^k) \leqslant f(x^k) + \lambda_k d_\phi(x^k, x^k) = f(x^k) \\
\Longrightarrow \quad & 0 \leqslant \lambda_k d_\phi(x^{k+1}, x^k) \leqslant f(x^k) - f(x^k + 1) \\
\Longrightarrow \quad & f(x^{k+1}) \leqslant f(x^k).
\end{aligned}$$

$\therefore$ La sucesión $\{f(x^k)\}_{k\in\mathbb{N}}$ es decreciente y además pues por hipótesis (A2) garantiza mínimo, así la sucesión está acotada y de acá, la sucesión es convergente, pues toda sucesión monótona y acotada converge.

(b) Por (3.5) y para todo $x \in U$ tenemos:

$$\begin{aligned}
& f(x^{k+1}) - f(x) \leqslant \lambda_k(d_\phi(x, x^k) - d_\phi(x^{k+1}, x^k)) \\
\Longrightarrow\ & 0 \leqslant \inf_{k\in\mathbb{N}} f(x^{k+1}) - f(x) \leqslant f(x^{k+1}) - f(x) \leqslant \lambda_k(d_\phi(x, x^k) - d_\phi(x^{k+1}, x^k)) \\
\Longrightarrow\ & 0 \leqslant \lambda_k(d_\phi(x, x^k) - d_\phi(x^{k+1}, x^k)) \\
\Longrightarrow\ & d_\phi(x, x^k) \leqslant d_\phi(x^{k+1}, x^k).
\end{aligned} \tag{3.6}$$

Como U es convexo, por la definición (2.16) se tiene que x^{k+1} es la única proyección de x^k en U con respecto a d_ϕ. Luego, por el lema (2.5):

$$\langle x - x^{k+1}, \nabla_1 d_\phi(x^{k+1}, x^k)\rangle \geqslant 0, \quad \forall x \in U, \tag{3.7}$$

así, aplicando el lema(2.6)(a) sustituyendo los puntos $c = x$, $a = x^k$ y $b = x^{k+1}$ se obtiene que:

$$H(x, x^k) - H(x, x^{k+1}) \geqslant \frac{\langle x - x^{k+1}, -\nabla_1 d_\phi(x^{k+1}, x^k)\rangle}{\mu\varphi''(1)\operatorname{máx}_{1\leqslant j\leqslant n}\{x_j^k\}}.$$

Como $\mu\varphi''(1)\operatorname*{máx}_{1\leqslant j\leqslant n}\{x_j^k\} > 0$ y usando 3.7 podemos concluir que:

$$H(x, x^k) \geqslant H(x, x^{k+1}).$$

(c) Por la parte anterior tenemos que la sucesión $\{H(x, x^k)\}_{k\in\mathbb{N}}$ es monótona y por la no negatividad de H tenemos que:

$$0 \leqslant H(x, x^k) \leqslant \ldots \leqslant H(x, x^0)$$

lo cual prueba que la sucesión $\{H(x, x^k)\}_{k\in\mathbb{N}}$ es monótona y acotada, así es convergente.

Lema 3.3. *Sea $\{\lambda_k\}_{k\in\mathbb{N}}$ una sucesión arbitraria de números positivos y $\{x^k\}_{k\in\mathbb{N}}$ la sucesión generada por RIPM. Entonces, bajo las hipótesis (A1) y (A2):*

(a) $\{f(x^k)\}_{k\in\mathbb{N}}$ es una sucesión decreciente y convergente.

(b) $\{x^k\}_{k\in\mathbb{N}}$ es Fejér convergente a el conjunto U.

(c) $\forall x \in U$, la sucesión $\{\|x - x^k\|\}_{k\in\mathbb{N}}$ es convergente.

Demostración.

(a) La demostración es exactamente igual la parte (a) del lema anterior, pues los terminos ν y μ estan implícitos en d_ϕ.

(b)Siguiendo la demostración del lema anterior podemos obtener las desigualdades (3.7) y (3.6). De esta manera, si utilizamos el lema 2.6 (b) y cambiando $a = x^k$, $b = x^{k+1}$ y $c = x$ tenemos:

$$0 \leqslant \langle x - x^{k+1}, \nabla_1 d_\phi(x^{k+1}, x^k) \rangle \leqslant \theta(\|x - x^k\|^2 - \|x - x^{k+1}\|^2),$$

donde $\theta = \dfrac{\nu + \mu\varphi''(1)}{2}$, luego bajo la hipótesis de que $\mu > 0$ se tiene que $\theta > 0$ y así:

$$0 \leqslant \theta(\|x - x^k\|^2 - \|x - x^{k+1}\|^2) \Longrightarrow \|x - x^k\|^2 \geqslant \|x - x^{k+1}\|^2, \quad \forall x \in U.$$

$\therefore \{x^k\}_{k \in \mathbb{N}}$ Fejér convergente.

(c) Usando la no negatividad de la $\|.\|$ y la parte (b) tenemos que:

$$0 \leqslant \|x - x^{k+1}\| \leqslant \|x - x^k\| \leqslant \ldots \leqslant \|x - x^0\|,$$

lo cual prueba que la sucesión es acotada, por la pate (b) es monótona y de esta manera converge.

Capítulo 4

Convergencia de los métodos

En este capítulo se probará la convergencia de los algoritmos (**IPM**) y (**RIPM**), bajo las hipótesis y argumentos enunciados en los capítulos anteriores.

Proposición 4.1. *Supongamos que las hipótesis (A1) y (A2) son satisfechas. Sea $\{\lambda_k\}_{k\in\mathbb{N}}$ una sucesión arbitraria de números positivos y $\{x^k\}_{k\in\mathbb{N}}$ la sucesión generada por **IPM**. Entonces, La sucesión $\{x^k\}_{k\in\mathbb{N}}$ converge, y además:*

(a) *Si existe $\underline{\lambda}$ y $\bar{\lambda}$ tal que $0 < \underline{\lambda} < \lambda_k < \bar{\lambda}$ para todo k, entonces:*

$$\lim_{k\to\infty} inf g_i^k \geqslant 0, \qquad \lim_{k\to\infty} g_i^k x_i^k = 0, \qquad \forall i = 1, \ldots, n, \tag{4.1}$$

donde $g^k \in \widehat{\partial} f(x^k)$ y g_i^k es la i-ésima componente de g^k.

(b) *Si $\lim_{k\to\infty} \lambda_k = 0$, entonces $\{x^k\}_{k\in\mathbb{N}}$ converge a una solución del problema* (5).

Demostración. Probemos en primer lugar que la sucesión $\{x^k\}_{k\in\mathbb{N}}$ converge. Por el lema 3.2(b), la sucesión $\{x^k\}_{k\in\mathbb{N}}$ es Fejér convergente a el conjunto U, con respecto a H y esto implica que:

$$\{x^k\}_{k\in\mathbb{N}} \subset \{y \in \mathbb{R}^n_{++} \,|\, H(x,y) \leqslant H(x,x_0)\}, \;\; \forall x \in U,$$

como una consecuencia del lema 2.3(a) la sucesión $\{x^k\}_{k\in\mathbb{N}}$ es acotada. Luego, existe $\bar{x}$ y una subsucesión $\{x^{k_j}\}_{k\in\mathbb{N}}$ de $\{x^k\}_{k\in\mathbb{N}}$ tal que $x^{k_j} \to \bar{x}$ (Toda sucesión acotada posee una subsucesión convergente). Luego por la semicontinuidad inferior de f:

$$\lim_{j\to\infty} f(x^{k_j}) \geqslant f(\bar{x}) \qquad\qquad (Pues\ f\ es\ convergente),$$

así, junto con el lema 3.2(a) (convergencia y decrecimiento de f) implica:

$$f(\bar{x}) \leqslant f(x^k),\ \ k \in \mathbb{N},$$

Con lo cual se prueba que $\bar{x} \in U$ por definición 3.1.

Ahora, por el lema 3.2(c) La sucesión $\{H(\bar{x}, x^k)\}_{k\in\mathbb{N}}$ converge y por el lema 2.3(b) tenemos que $\lim_{j\to\infty} H(\bar{x}, x^{k_j}) = 0$.

Por tanto, $\lim\limits_{k\to\infty} H(\bar{x}, x^k) = 0$.

Luego por el Lema 2.3(c), cambiando $z^k = x^k$,$y^k = \bar{x}$ y como $\lim\limits_{k\to\infty} H(\bar{x}, x^k) = 0$, se tiene asi que $\{x^k\}_{k\in\mathbb{N}}$ converge a $\bar{x}$.

Probemos (a). Por el esquema iterativo (6) x^{k+1} es solución y por el lema 2.1(d):

$$0 \in \widehat{\partial}(f(x) + \lambda_k d_\phi(x, x^k)(x^{k+1})),$$

así, por le lema 2.1(c) tenemos que existe $g^{k+1} \in \widehat{\partial} f(x^{k+1})$ tal que:

$$g^{k+1} + \lambda_k \nabla_1 d(x^{k+1}, x^k) = 0 \Longleftrightarrow -g^{k+1} = \lambda_k \nabla_1 d(x^{k+1}, x^k). \tag{4.2}$$

Por otra parte, sabemos que:

$$\nabla_1 d(x_i^{k+1}, x_i^k) = \mu x_i^k \varphi'(\frac{x_i^{k+1}}{x_i^k}),$$

asi, la ecuación 4.2 nos queda como:

$$-\,g^{k+1} = \lambda_k \mu x_i^k \varphi'(\frac{x_i^{k+1}}{x_i^k}),\ \ \forall i = 1, \ldots, n. \tag{4.3}$$

Definamos los siguientes conjuntos de índices:

$$I(\bar{x}) = \{i \in \{1, 2, \ldots, n\} \mid \bar{x}_i > 0\}\ \ y\ \ J(\bar{x}) = \{i \in \{1, 2, \ldots, n\} \mid \bar{x}_i = 0\}\,.$$

Estudiemos ahora los siguientes casos:

Caso(i): $i \in I(\bar{x})$. En este caso:

$$\lim_{k \longrightarrow \infty} \frac{x_i^{k+1}}{x_i^k} = 1.$$

Es válido pues $x^k \longrightarrow \bar{x}$, usando la continuidad de φ', el hecho de que $\varphi'(1) = 0$ y además que λ_k es acotada, al aplicar límite a la igualdad (4.3) tenemos:

$$\lim_{k \longrightarrow \infty} g^{k+1} = 0, \quad \forall i \in I(\bar{x}) \tag{4.4}$$

Caso(ii): $i \in J(\bar{x})$. Para este se definen los siguientes conjuntos de índices:

$$J_+^i = \left\{ k : \frac{x_i^{k+1}}{x_i^k} > 1 \right\} \;\; y \;\; J_-^i = \left\{ k : \frac{x_i^{k+1}}{x_i^k} \leqslant 1 \right\}.$$

Estudiemos J_+^i. Como φ' es monótona creciente en este dominio tenemos:

$$\begin{aligned}
\frac{x_i^{k+1}}{x_i^k} > 1 &\Longrightarrow \varphi'(1) \leqslant \varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) \\
&\Longrightarrow \varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) \geqslant 0 \\
&\Longrightarrow \mu\lambda_k\varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) \geqslant 0 \\
&\Longrightarrow -g_i^{k+1} \geqslant 0 \\
&\Longrightarrow g_i^{k+1} \leqslant 0, \qquad\qquad \forall k \in J_+^i,\ \forall i \in J(\bar{x}).
\end{aligned} \tag{4.5}$$

Por otra parte, usando (4.3) y la hipótesis $\varphi \in \Phi_1 \subset \Phi_2$ tenemos:

$$\begin{aligned}
g^{k+1} = -\lambda_k\mu x_i^k\varphi'(\frac{x_i^{k+1}}{x_i^k}) &\geqslant -\lambda_k\mu x_i^k\varphi''(1)(\frac{x_i^{k+1}}{x_i^k} - 1) \\
&= -\lambda_k\mu\varphi''(1)(x_i^{k+1} - x_i^k) \\
&\geqslant -\overline{\lambda}\mu\varphi''(1)(x_i^{k+1} - x_i^k) \quad \forall k \in J_+^i.
\end{aligned} \tag{4.6}$$

Note además que por la convergencia de $\{x^k\}_{k\in\mathbb{N}}$, $\lim_{k\longrightarrow\infty}(x^{k+1} - x^k) = 0$. Así, aplicando límite a (4.6) tenemos:

$$\lim_{k \longrightarrow \infty} g_i^{k+1} \geqslant 0. \tag{4.7}$$

Asi, de (4.5) y (4.7) podemos concluir que:

$$\lim_{k\longrightarrow\infty,\ k\in J_+^i} g_i^{k+1} = 0, \qquad \forall i \in J_+^i(\bar{x}). \tag{4.8}$$

Para la otra parte, $\varphi'(t) \leqslant 0$ para todo $0 < t \leqslant 1$, usando esto en (4.3) entonces para J_-^i,

$$g_i^{k+1} \geqslant 0 \ \ \forall k \in J_-^i, \forall i \in J(\bar{x}),$$

luego con esta última igualdad queda probado que:

$$\lim_{k\to\infty} \inf g_i^k \geqslant 0. \tag{4.9}$$

Por último, Veamos que $\lim\limits_{k\to\infty} g_i^k x_i^k = 0$. Usando (4.8), (4.4) y el hecho de que $\{x^k\}_{k\in\mathbb{N}}$ converge a $\bar{x}$, solo se tiene que probar que:

$$\lim_{k\longrightarrow\infty,\ k\in J_-^i} g_i^{k+1} x_i^{k+1} = 0, \quad \forall i \in J(\bar{x}).$$

Por definición:

$$\lim_{k\longrightarrow\infty,\ k\in J_-^i} x_i^{k+1} = 0, \forall i \in\ J(\bar{x}),$$

y además, por (4.9) solo se debe probar que la sucesión $\{g_i^k\}_{k\in J_-^i}$ para cada $i \in J(\bar{x})$ está acotada superiormente. Así, tomemos $\varepsilon_0 > 0$ y $x \in \mathbb{R}_{++}^n \cap dom(f)$ con $x_i > \varepsilon_0$ para cada i. Entonces para $k \in J_-^i$ suficientemente grande, se tiene:

$$x_i - x_i^{k+1} \geqslant \frac{\varepsilon_0}{2}, \quad \forall i \in J(\bar{x}). \tag{4.10}$$

Por la definición (2.7) tenemos:

$$f(x) \geqslant f(x^{k+1}) + \sum_{i=1}^{n} g_i^{k+1}(x_i - x_i^{k+1}) + \circ(\|x - x^{k+1}\|); \tag{4.11}$$

esto implica que la sucesión $\{g_i^{k+1}\}_{k\in J_-^i}$ es acotada para cada $i \in J(\bar{x})$. Pues, si suponemos lo contrario, es decir que existe un índice $i_0 \in J(\bar{x})$ y una subsucesión $\{g_{i_0}^{k_l}\}_{k_l\in J_-^i}$ (*Con* $\lim\limits_{l\to\infty} k_l = +\infty$) tal que:

$$\lim_{l\longrightarrow+\infty} g_{i_0}^{k_l+1} = +\infty, \ \ g_{i_0}^{k_l+1} \geqslant 0.$$

Luego, como la sucesión $\{x^{k+1}\}_{k\in\mathbb{N}}$ es convergente, usando las ecuaciones (4.8)-(4.11) tenemos que existe un $\eta \in \mathbb{R}$ tal que para un l suficientemente grande, donde los terminos $i \neq i_0$ y por ende los téminos $g_i^{k_l+1}$ adquieren valores finitos se obtiene:

$$\sum_{i\neq i_0} g_i^{k_l+1}(x_i - x_i^{k_l+1}) + \circ(\|x - x^{k_l+1}\|) \geqslant \eta,$$

y por las ecuaciones (4.10) y (4.11), obtenemos que

$$f(x) \geqslant f(x^{k_l+1}) + \frac{\varepsilon_0}{2} g_{i_0}^{k_l+1} + \eta,$$

Luego, $\lim\limits_{l\to+\infty} f(x^{k_l+1}) \geqslant f(\bar{x})$ (pues f es inferiormente semicontinua y convergente), asi aplicando límite a la expresión anterior tenemos que $f(x) \geqslant \infty$ lo cual es una contradicción, pues $x \in dom(f)$.

(b) Por la desigualdad (3.5) y la no negatividad de d_ϕ, se obtiene que:

$$f(x^{k+1}) \leqslant f(x) + \lambda_k d_\phi(x, x^k), \quad \forall x \in \mathbb{R}^n_{++},$$

tomando límite cuando $k \to +\infty$ sobre la desigualdad anterior y usando la hipótesis de que $\lim\limits_{k\to+\infty} \lambda_k = 0$, el lema (2.4)(e) y la semicontinuidad inferior de f se obtiene que:

$$f(\bar{x}) \leqslant f(x), \forall x \in \mathbb{R}^n_{++},$$

donde $\bar{x}$ es tal que $\lim\limits_{k\to+\infty} x^k = \bar{x}$. Esto implica que $\bar{x} \in X^*$.

■

PROPOSICIÓN 4.2. *Supongamos que las hipótesis (A1) y (A2) son satisfechas. Sea $\{\lambda_k\}_{k\in\mathbb{N}}$ una sucesión arbitraria de números positivos y $\{x^k\}_{k\in\mathbb{N}}$ la sucesión generada por **RIPM**. Entonces, La sucesión $\{x^k\}_{k\in\mathbb{N}}$ converge, y además:*

(a) *Si existe $\underline{\lambda}$ y $\bar{\lambda}$ tal que $0 < \underline{\lambda} < \lambda_k < \bar{\lambda}$ para todo k, entonces:*

$$\lim_{k\to\infty} inf g_i^k \geqslant 0, \qquad \lim_{k\to\infty} g_i^k x_i^k = 0, \qquad \forall i = 1, \ldots, n, \tag{4.12}$$

donde $g^k \in \widehat{\partial} f(x^k)$ y g_i^k es la i-ésima componente de g^k.

(b) *Si $\lim\limits_{k\to\infty} \lambda_k = 0$, entonces $\{x^k\}_{k\in\mathbb{N}}$ converge a una solución del problema (5).*

Demostración. Veamos que la sucesión $\{x^k\}_{k\in\mathbb{N}}$ converge. Por el lema 3.3(b), la sucesión $\{x^k\}_{k\in\mathbb{N}}$ es fejér convergente a el conjunto U, lo cual implica que:

$$\{x^k\}_{k\in\mathbb{N}} \subset \{y \in \mathbb{R}^n_{++} \ : \|x-y\| \leqslant \|x-x_0\|\} \ \ \forall x \in U.$$

Observemos que el conjunto de la derecha es acotado, para todo $x \in \mathbb{R}^n$ y así, la sucesión $\{x^k\}_{k\in\mathbb{N}}$ es acotada, por tanto existe una subsucesion $\{x^{k_j}\}_{k_j} \in \mathbb{N}$ de $\{x^k\}_{k\in\mathbb{N}}$ que converge a $\widehat{x}$. Usando un argumento similar al de la proposición anterior se tiene que $\widehat{x} \in U$. Luego, por el lema 3.3(c), la sucesión $\{\|x^k - \widehat{x}\|\}_{k\in\mathbb{N}}$ es convergente. En consecuencia, $\{x^{k_j}\} \in \mathbb{R}^n_{++}$ converge a $\widehat{x} \in \mathbb{R}^n_{+}$. Por otra parte, por la continudad de la norma, $\|x^{k_j} - \widehat{x}\| \to 0$ y en consecuencia, $\|x^k - \widehat{x}\| \to 0$, esto implica que $x^k \to \widehat{x}$.

(a) Por la fórmula iterativa (6) y el lema 2.1(d) se tiene:

$$0 \in \widehat{\partial}(f(x) + \lambda_k d_\phi(x, x^k))(x^{k+1}),$$

de esta manera, existe $h^{k+1} \in \widehat{\partial} f(x^{k+1})$ tal que:

$$\lambda_k \nabla_1 d_\phi(x^{k+1}, x^k) = -h^{k+1}.$$

Ahora, sustituyendo el gradiente de d_ϕ, que para cada $i = 1, \ldots, n$

$$h^{k+1} = -\mu\lambda_k x_i^k \varphi'(\frac{x_i^{k+1}}{x_i^k}) - \nu\lambda_k(x_i^{k+1} - x_i^k). \tag{4.13}$$

Luego, para $\varphi \in \Phi_2$ se tiene por (2.33) que para cada $i = 1, \ldots, n$

$$-\mu\lambda_k x_i^k \varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) \geqslant -\mu\lambda_k \varphi''(1) x_i^k \left(\frac{x_i^{k+1}}{x_i^k} - 1\right) = -\mu\lambda_k\, \varphi''(1)(x_i^{k+1} - x_i^k),$$

restando $\nu\lambda_k(x_i^{k+1} - x_i^k)$ a la última desigualdad tenemos:

$$h_i^{k+1} \geqslant \lambda_k(\mu\varphi''(1) + \nu)(x_i^k - x_i^{k+1}) = 2\theta\lambda_k(x_i^k - x_i^{k+1}), \quad i = 1, \ldots, n. \tag{4.14}$$

Asi considerando que la sucesión $\{x^k\}_{k\in\mathbb{N}}$ es convergente y además que λ_k es acotada, tomando límite cuando $k \to \infty$ en (4.14) se tiene la primera parte de (4.12).

Para la segunda parte, vamos a tomar en cuenta el lado izquierdo de la desigualdad con la cual se definen las clases de funciones $\varphi \in \Phi_2$, se obtiene:

$$-\mu\lambda_k x_i^k \varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) \leqslant \mu\lambda_k\varphi''(1)(x_i^k - x_i^{k+1})\left(\frac{x_i^k}{x_i^{k+1}}\right), \quad i = 1, \ldots, n.$$

Ahora, restando $\nu\lambda_k(x_i^{k+1} - x_i^k)$ a la desigualdad anterior tenemos:

$$h_i^{k+1} \leqslant \mu\lambda_k\varphi''(1)(x_i^k - x_i^{k+1})\left(\frac{x_i^k}{x_i^{k+1}}\right) + \nu\lambda_k(x_i^k - x_i^{k+1}), \quad i = 1, \ldots, n. \tag{4.15}$$

Combinando las desigualdades (4.14) y (4.15) tenemos:

$$2\theta\lambda_k(x_i^k - x_i^{k+1}) \leqslant h_i^{k+1} \leqslant (\mu\lambda_k\varphi''(1)\left(\frac{x_i^k}{x_i^{k+1}}\right) + \nu\lambda_k)(x_i^k - x_i^{k+1}).$$

Asi, multiplicando por x_i^{k+1} nos queda:

$$2\theta\lambda_k x_i^{k+1}(x_i^k - x_i^{k+1}) \leqslant h_i^{k+1}x_i^{k+1} \leqslant (\mu\lambda_k\varphi''(1)x_i^k + \nu\lambda_k x_i^{k+1})(x_i^k - x_i^{k+1}).$$

Tomando límite cuando $k \to \infty$, considerando que $\{x_i^k\}_{k\in\mathbb{N}}$ es convergente y que λ_k es acotada, se tiene probada la parte (a).

(b) La demostración es exactamente igual a la parte (b) de la proposición anterior, (tomando en cuenta que ν y μ cumplen con las hipótesis enunciadas en esta proposición) ya que las variables ν y μ estan implicitas en d_ϕ.

■

Corolario 4.1. *Suponga que f es una función cuasi-convexa, continuamente diferenciable en $\mathbb{R}^n$ y supongamos que las hipótesis (A1) y (A2) son satisfechas. Sea $\{x^k\}_{k\in\mathbb{N}}$ la sucesión generada por* ***IPM*** *o* ***RIPM*** *entonces:*

(a) *Si existen $\underline{\lambda}$ y $\bar{\lambda}$ tales que, $0 < \underline{\lambda} < \lambda_k < \bar{\lambda}$ para todo k, entonces $\{x^k\}_{k\in\mathbb{N}}$ converge a un punto estacionario del problema (6).*

(b) *Si $\lim\limits_{k\to\infty} \lambda_k = 0$, entonces $\{x^k\}_{k\in\mathbb{N}}$ converge a una solución del problema* (6).

Demostración.

(a) Por las proposiciones (4.1),(4.2), usando el hecho que sucesión $\{x^k\}_{k\in\mathbb{N}}$ es convergente y denotamos su límite por $\bar{x}$. Entonces, usando el Lema 2.1(b) y (4.1) ó (4.12) tenemos:

$$(\nabla f(\bar{x})_i)\bar{x}_i = \lim_{k\to\infty}(\nabla f(x^k))_i x_i^k = 0, \quad i = 1, \ldots, n,$$

esto implica que $\nabla f(\bar{x})_T(x-\bar{x}) \geqslant 0$, para todo $x \geqslant 0$. Por tanto, por definición (2.8)(a) $\bar{x}$ es un punto estacionario del problema (5).

(b) La demostración es análoga a la parte (b) de la proposición (4.1) ó (4.2) y debido a esto se omite. ∎

DEFINICIÓN 4.1. *Sea $\pi_U : \mathbb{R}^n \to \mathbb{R}^n$ tal que:*

$$\pi_U(x^0) = argmin\{\|x - x^0\| \ : \ x \in U\};$$

se define:

$$\rho(x^0, U) = \|\pi_U(x^0) - x^0\|.$$

PROPOSICIÓN 4.3. *Supongamos que las hipótesis (A1) y (A2) son satisfechas. Sea $\{\lambda_k\}_{k\in\mathbb{N}}$ una sucesión de números positivos y $\{x^k\}_{k\in\mathbb{N}}$ la sucesión generada por* ***RIPM*** *con límite $\bar{x}$. Entonces,*

$$\|x^0 - \bar{x}\| \leqslant 2\rho(x^0, U).$$

Demostración. Como la sucesión $\{x^k\}_{k\in\mathbb{N}}$ es fejér convergente entonces si tomamos $x = \pi_U(x^0)$ obtenemos que $\|\pi_U(x^0) - x^k\| \leqslant \|\pi_U(x^0) - x^0\|$. Así, tomando límite cuando $k \to \infty$ se tiene que:

$$\|\pi_U(x^0) - \bar{x}\| \leqslant \|\pi_U(x^0) - x^0\|.$$

con lo cual se obtiene que:

$$
\begin{aligned}
\|x^0 - \bar{x}\|^2 &\leqslant (\|x^0 - \pi_U(x^0)\| + \|\bar{x} - \pi_U(x^0)\|)^2 \\
&\leqslant 2(\|x^0 - \pi_U(x^0)\|^2 + \|\bar{x} - \pi_U(x^0)\|^2) \\
&\leqslant 2(\|x^0 - \pi_U(x^0)\|^2 + \|x^0 - \pi_U(x^0)\|^2) \\
&\leqslant 4\|x^0 - \pi_U(x^0)\|^2 \\
&= 4\rho^2(x^0, U).
\end{aligned}
$$

Por último, tomando raíz cuadrada tenemos:

$$\|x^0 - \bar{x}\| \leqslant 2\rho(x^0, U),$$

Con lo cual, la prueba está completa. ■

Capítulo 5

Conclusiones

En este trabajo, se desarrollaron dos clases de algoritmos proximales, basados en una casi distancia homogénea de segundo orden, llamados (IPM) y (RIPM) para el problema de optimización cuasi-convexa definido en (5). Las propiedades de convergencia de los algoritmos fueron establecidas bajo algunas hipótesis y particularmente se obtuvieron dos resultados de convergencia importantes, debido a la imposición de ciertas condiciones en los parámetros proximales λ_k, (vease 4.1 y 4.2) y en particular una consecuencia en el caso donde la función a minimizar sea continua.

Referencias bibliográficas.

[1] Shaohua Pan and Jein-Shan Chen. *Entropy-like proximal algorithms based on a second-order homogeneous distance function for quasi-convex programming.*Journal of Global Optimization, vol. 39, pp. 555-575, 2007.

[2] M. S. Bazaraa and C. M. Shetty. *Nonlinear Programming Theory and Algorithms* New York: John Wiley y Sons, 1979.

[3] A. Iusem, B. Svaiter, and M. Teboulle. *Entropy-Like proximal methods in convex programming, Mathematics of Operation Research.*Mathematics of Operation Re-search, vol. 20. 1995.

[4] R. T. Rockafella. *Convex Analysis.*Princeton University Press. Princeton New Jersey,1970.

[5] R.T. Rockafellar and R. Wets. *Variational analysis.* Springer-Verlag. 1998.

[6] Jean-Baptiste Hiriart-Urruty, Claude Lemaréchal. *Convex analysis and minimization algorithms.* Springer. 1996.