UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas

“Un algoritmo de punto Proximal que involucra una función objetivo no-convexa. ”

Trabajo Especial de Grado presentado por

Br Luis Freytez.

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

Barquisimeto, Venezuela. Junio de 2009

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:

"UN ALGORITMO DE PUNTO PROXIMAL QUE INVOLUCRA UNA FUNCIÓN OBJETIVO NO-CONVEXA. "

presentado por el ciudadano BR LUIS FREYTEZ. titular de la Cédula de Identidad No. 16.794.821, 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

A Genis y Marlene

Agradecimientos

Primeramente quiero agradecer a DIOS por darme la vida y guiar mis pasos sin pedir nada a cambio...

A mi papa y mi mama por su apoyo en todo momento en todos los aspectos, sin ellos no seria ni la mitad de lo que soy. Gracias....

A mis hermanas por estar siempre allí y tener dos amigas con quien puedo conversar...

A todos los Profesores que compartieron gran parte de sus conocimientos conmigo...

A el DR. Rómulo Castillo persona que admiro mucho y que hizo posible el desarrollo de este trabajo...

A mis compañeros de estudio que estuvimos juntos desde el comienzo de la carrera hasta el final (Maria, Toñeco, Tony, Joel y Marilyn) nunca olvidare todas esas experiencias que vivimos juntos...

A mis compañeros Licenciados de la promoción 39 en especial a (Elifer(El mardito), Adrian, Manuel y Yuly)...

A mi grupo de esparcimiento el bien llamado los anormales (David, Manuel, Mario, Memo) y los agregados (Adrian, Elifer ,Gregorio y Marco) andando juntos en aventuras trasladados por el Poncho azul, casi siempre para que el Portu...

A Katherine, mi novia, que si bien la conocí fue al final de mi carrera no importa se que valió la pena esperar a alguien tan valiosa como ella, también contribuyó a que el ultimo semestre de mi carrera fuera el mejor...

A todo el que contribuyó en mi desarrollo tanto como persona y profesional y no

nombre también le agradezco...

Resumen

En este trabajo, estudiaremos un algoritmo casi-proximal para minimizar una función f cerrada y propia sujeto a restricciones de no-negatividad $x \geq 0$, basado en el esquema iterativo

$$x^k = argmin_{x \geq 0}\{f(x) + \mu_k d(x, x^{k-1})\},$$

donde $d(\cdot, \cdot)$ es una casi-distancia entrópica. El algoritmo esta bien definido bajo la suposición que el problema tiene un conjunto solución no-vacío y acotado. En adición, si f es una función diferenciable y casi-convexa, mostraremos que la sucesión generada por el algoritmo es convergente y además esta converge a una solución optima del problema cuando el parámetro μ_k se aproxima a 0.

Índice

Introducción

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

$$\begin{aligned} x^0 &\in \mathbb{R}^n, \\ x^k &= argmin_{x\in\mathbb{R}^n}\{f(x) + \mu_k\|x - x^{k-1}\|^2\}, \end{aligned} \tag{1}$$

donde μ_k es una sucesión de números positivos y $\|.\|$ denota la norma Euclidiana en $\mathbb{R}^n$.

El algoritmo de punto proximal fue inicialmente estudiado y desarrollado por Rockafellar [6],[7]. En 1992, Teboulle [9] modifica el método de punto proximal reemplazando la distancia cuadrática en (1) por la siguiente casi-distancia, también conocida con el nombre de φ-divergencia:

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

donde $\varphi : \mathbb{R}_{++} \to \mathbb{R}$ es una función propia y cerrada que satisface ciertas condiciones (ver [8]). Una importante elección de φ es $\varphi(t) = tlnt - t + 1$, para la cual la correspondiente d_φ es bien conocida como Kullback-Leibler [8] función entrópica de estadísticas.

El algoritmo asociado con la φ-divergencia para minimizar una función convexa f sujeto a restricciones de no-negatividad $x \geq 0$ esta dado como sigue:

$$\begin{aligned} x^0 &> 0, \\ x^k &= argmin_{x\geq 0}\{f(x) + \mu_k d_\varphi(x, x^{k-1})\}, \end{aligned} \tag{3}$$

donde μ_k es una sucesión de números positivos . El algoritmo en (3) es llamado *casi-proximal*, este algoritmo con $\varphi(t) = -lnt + t - 1$ fue propuesto primero por Eggermont en [4]. Un punto importante de la diferencia entre (1) y (3) es que d_φ es usada para forzar a los iterados $\{x^k\}_{k\in\mathbb{N}}$ a quedarse en el ortante no-negativo $\mathbb{R}^n_+$, es decir, el algoritmo en (3) genera automáticamente una sucesión de términos positivos $\{x^k\}_{k\in\mathbb{N}}$. Similares resultados de convergencia del método casi-proximal se obtienen usando distancias de Bregman(ver [3]). El análisis del método casi-proximal basado en distancias de Bregman no puede ser llevado a cabo igual que en el algoritmo definido en (3), excepto en el caso que $\varphi(t) = tlnt - t + 1$ donde las dos distancias coinciden. También podemos observar que el algoritmo en (3) fue adoptado por [2] para resolver problemas de minimizar una función cerrada f sobre $\mathbb{R}^n$ sin asumir la convexidad de f.

En este trabajo estudiaremos el algoritmo definido en (3) con

$$\varphi(t) = -lnt + t - 1 \ \ (t > 0), \tag{4}$$

para resolver el problema de minimización no-convexo de la siguiente forma:

$$\begin{aligned} &min\ f(x) \\ &s.t\ \ x \geq 0, \end{aligned} \tag{5}$$

donde $f : \mathbb{R}^n \to \mathbb{R}$ es una función propia cerrada con dom$f \subseteq \mathbb{R}^n_{++}$. Como no requerimos la convexidad de f, el algoritmo en este trabajo es el siguiente:

$$\begin{aligned} x^0 &> 0, \\ x^k &\in argmin_{x\geq 0}\{f(x) + \mu_k d(x, x^{k-1})\}, \end{aligned} \tag{6}$$

donde μ_k es una sucesión de números positivos y $d(x, y)$ es especificada como sigue:

$$d(x,y) = \sum_{i=1}^{n} y_i\varphi(x_i/y_i) = \sum_{i=1}^{n}[y_i ln(y_i/x_i) + x_i - y_i]. \tag{7}$$

El principal propósito en este trabajo es establecer los resultados de convergencia del algoritmo (6)-(7) bajo algunas suposiciones del problema (5).

Capítulo 1
Preliminares

1.1. Conceptos Básicos

En esta sección, presentaremos algunos resultados preliminares que serán usados en los siguientes capítulos.

Definición 1.1. Sea $\varphi : \mathbb{R}_{++} \to \mathbb{R}$ convexa y tres veces diferenciable, que satisface

$$\varphi(1) = \varphi'(1) = 0, \quad \varphi''(1) > 0 \quad \text{y} \quad \lim_{t \to 0} \varphi'(t) = -\infty,$$

entonces $d_\varphi : \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \longrightarrow \mathbb{R}$ dada por

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

es llamada φ-*divergencia*.

Definición 1.2. Una sucesión $\{x^k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n_{++}$ es Fejér convergente al conjunto no vacío $\mathcal{U} \subseteq \mathbb{R}^n$ con respecto a la divergencia $d(\cdot, \cdot)$, si $d(x^k, u) \leq d(x^{k-1}, u)$ para cada $k \in \mathbb{N}$ y $u \in \mathcal{U}$.

Definición 1.3. Dada una función de valor real extendida $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, el dominio efectivo de f se define como: $dom(f) = \{x \in \mathbb{R}^n : f(x) < +\infty\}$.

Definición 1.4. Dada una función de valor real extendida $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, el epigrafo de f es el conjunto: $epi(f) = \{(x, \beta) \in \mathbb{R}^n \text{ x } \mathbb{R} : f(x) \leq \beta\}$.

Definición 1.5. Una función f es propia si $dom(f) \neq \varnothing$ y $f(x) > -\infty$ para todo $x \in dom(f)$.

Definición 1.6. Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ es llamada cerrada si es semicontinua inferiormente o si su epigrafo es un conjunto cerrado o si sus conjuntos de nivel son cerrados.

Definición 1.7. Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función propia. Entonces,

(a) f es llamada *casi-convexa* si para todo $x, y \in dom(f)$ y $\alpha \in (0, 1)$

$$f(\alpha x + (1-\alpha)y) \leq max\{f(x), f(y)\}.$$

(b) f es llamada *estrictamente casi-convexa* si para todo $x, y \in dom(f)$ con $f(x) \neq f(y)$

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

Lema 1.1. *(ver[1]) Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función propia. Entonces,*

(a) *f es casi-convexa si y sólo si los conjuntos de nivel $L_f(\gamma)$ dado por $L_f(\gamma) = \{x \in dom(f) : f(x) \leq \gamma\}$ son convexos para todo $\gamma \in \mathbb{R}$.*

(b) *f es diferenciable en $dom(f)$, entonces f es casi-convexa si y solo si $\langle \nabla f(x_2), x_1 - x_2 \rangle \leq 0$ siempre que $f(x_1) \leq f(x_2)$ para todo $x_1, x_2 \in dom(f)$.*

Demostración. (a)($\Rightarrow$) Supongamos que f es casi convexa y sean $x_1, x_2 \in L_f(\gamma)$. Como $x_1, x_2 \in L_f(\gamma)$ se tiene que $x_1, x_2 \in dom(f)$ y $max\{f(x_1), f(x_2)\} \leq \gamma$, para algún $\gamma \in \mathbb{R}$. Sea $\lambda \in (0, 1)$ y $x = \lambda x_1 + (1-\lambda)x_2$, puesto que $dom(f)$ es convexo $x \in dom(f)$ además por la casi-convexidad de f tenemos que

$$f(x) = f(\lambda x_1 + (1-\lambda)x_2) \leq max\{f(x_1), f(x_2)\} \leq \gamma,$$

así $x \in L_f(\gamma)$ y por tanto $L_f(\gamma)$ es convexo.

($\Leftarrow$)Supongamos que $L_f(\gamma)$ es convexo para cada $\gamma \in \mathbb{R}$. Sean $x_1, x_2 \in dom(f)$ y $x = \lambda x_1 + (1-\lambda)x_2$ con $\lambda \in (0, 1)$. Notemos que $x_1, x_2 \in L_f(\gamma)$ para $\gamma = max\{f(x_1), f(x_2)\}$. Ahora por hipótesis, como $L_f(\gamma)$ es convexo entonces $x \in L_f(\gamma)$, por lo que $f(x) \leq \gamma = max\{f(x_1), f(x_2)\}$. Así f es casi-convexa.

(b) Supongamos que f es diferenciable en $dom(f)$.

($\Rightarrow$) Sea f casi convexa y $x_1, x_2 \in dom(f)$ tales que $f(x_1) \leq f(x_2)$. Por la diferenciabilidad de f en x_2 para $\lambda \in (0,1)$, tenemos

$$f[\lambda x_1+(1-\lambda)x_2]-f(x_2) = \lambda\nabla f(x_2)^t(x_1-x_2)+\lambda\|x_1-x_2\|\alpha[x_2;\lambda x_1+(1-\lambda)x_2] \quad (1.1)$$

donde $\alpha[x_2; \lambda x_1 + (1-\lambda)x_2] \to 0$ cuando $\lambda \to 0$. Como f es casi-convexa se cumple que $f[\lambda x_1 + (1-\lambda)x_2] \leq f(x_2)$, luego de la ecuación (1.1) tenemos que

$$\lambda\nabla f(x_2)^t(x_1 - x_2) + \lambda\|x_1 - x_2\|\alpha[x_2; \lambda x_1 + (1-\lambda)x_2] \leq 0$$

dividiendo por λ y $\lambda \to 0$, obtenemos que $\nabla f(x_2)^t(x_1 - x_2) \leq 0$.

($\Leftarrow$) Sean $x_1, x_2 \in dom(f)$ tales que $f(x_1) \leq f(x_2)$. Queremos probar que

$$f[\lambda x_1 + (1-\lambda)x_2] \leq f(x_2) \quad \text{para cada} \quad \lambda \in (0,1).$$

Lo haremos mostrando que

$$L = \{x : x = \lambda x_1 + (1-\lambda)x_2, \lambda \in (0,1), f(x) > f(x_2)\}$$

es vacío. Por el método de reducción al absurdo, supongamos que existe un $x^{'} \in L$. Por tanto $x' = \lambda x_1 + (1-\lambda)x_2$, para algún $\lambda \in (0,1)$ y $f(x') > f(x_2)$. Como f es diferenciable, entonces es continua, por lo que debe existir un $\delta \in (0,1)$ tal que

$$f[\mu x' + (1-\mu)x_2] > f(x_2) \qquad \forall \mu \in [\delta, 1] \quad (1.2)$$

y $f(x') > f[\delta x' + (1-\delta)x_2]$. Por esta última desigualdad y el Teorema del valor medio tenemos

$$0 < f(x') - f[\delta x' + (1-\delta)x_2] = (1-\delta)\nabla f(\hat{x})^t(x' - x_2) \quad (1.3)$$

donde $\hat{x} = \hat{\mu}x' + (1-\hat{\mu})x_2$, para algún $\hat{\mu} \in (\delta, 1)$. Usando (1.2) obtenemos que $f(\hat{x}) > f(x_2)$. Dividiendo (1.3) por $1-\delta > 0$, resulta $\nabla f(\hat{x})^t(x' - x_2) > 0$, lo cual implica que

$$\nabla f(\hat{x})^t(x_1 - x_2) > 0. \quad (1.4)$$

Pero por otro lado, $f(\hat{x}) > f(x_2) > f(x_1)$ con $\hat{x}$ una combinación convexa de x_1 y x_2 digamos $\hat{x} = \hat{\lambda}x_1 + (1-\hat{\lambda})x_2$, donde $\hat{\lambda} \in (0,1)$. Por hipótesis $\nabla f(\hat{x})^t(x_1 - \hat{x}) \leq 0$, de esto tenemos

$$0 \geq \nabla f(\hat{x})^t(x_1 - \hat{x}) = (1-\lambda)\nabla f(\hat{x})^t(x_1 - x_2),$$

así,

$$\nabla f(\hat{x})^t(x_1 - x_2) \leq 0$$

lo cual es una contradicción con (1.4). Por tanto L es vacío y así la demostración esta completa. $\square$

Definición 1.8. Para un problema de optimización restringida de minimizar f sobre $x \in C$ donde C es no vacío y es un subconjunto convexo de $\mathbb{R}^n$, x^* es llamado un punto estacionario si $\langle \nabla f(x^*), x - x^* \rangle \geq 0$.

Definición 1.9. Sea $S \subset \mathbb{R}^n$ abierto y convexo y $\bar{S}$ su clausura. Consideremos una función real convexa h definida en $\bar{S}$ y sea $D_h : \bar{S} \times S \to \mathbb{R}$ dada por

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

Entonces h es llamada función de Bregman (y D_h distancia de Bregman inducida por h) si se cumplen las siguientes condiciones:

(a) h es continuamente diferenciable en S.

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

(c) Para todo $\delta \in \mathbb{R}$ los conjuntos de nivel $L_1(y, \delta) = \{x \in \bar{S} : D_h(x, y) \leq \delta\}$ y $L_2(x, \delta) = \{y \in S : D_h(x, y) \leq \delta\}$ son acotados para todo $y \in$ S y $x \in \bar{S}$ respectivamente.

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

(e) Si $\{x^k\} \subset \bar{S}$ y $\{y^k\} \subset S$ son sucesiones tales que $\{x^k\}$ es acotada, $\lim\limits_{k \to +\infty} y^k = y^*$ y $\lim\limits_{k \to +\infty} D_h(x^k, y^k) = 0$, entonces $\lim\limits_{k \to +\infty} x^k = y^*$.

En lo que sigue nos concentraremos en las propiedades de φ dadas en (4) y la introducción de la función d , las cuales serán usadas para el posterior análisis. Primero haremos un resumen de las propiedades de φ.

1.2. Propiedades de la función φ

Sea $\varphi : \mathbb{R}_{++} \to \mathbb{R}$ definida como en (4). Entonces se cumplen los siguientes resultados:

(a) $\varphi(t) \geq 0$ y $\varphi(t) = 0$ si y solo si t $= 1$.

(b) $\varphi(t)$ es decreciente en (0,1) con $\lim\limits_{t\to 0^+} \varphi(t) = +\infty$ y creciente en $(1, \infty)$ con $\lim\limits_{t\to +\infty} \varphi(t) = +\infty$.

(c) $\varphi(1) = 0$, $\varphi^{'}(1) = 0$ y $\varphi^{''}(1) > 0$.

(d) $\varphi^{'}(t)$ es no decreciente en $(0, +\infty)$ y $\lim\limits_{t\to +\infty} \varphi^{'}(t) = 1$, $\lim\limits_{t\to 0^+} \varphi^{'}(t) = -\infty$.

(e) $\varphi^{'}(t) \leq \ln t$ para todo $t > 0$ y $\varphi^{'}(t) > 0$ cuando $t \in [1, \infty)$.

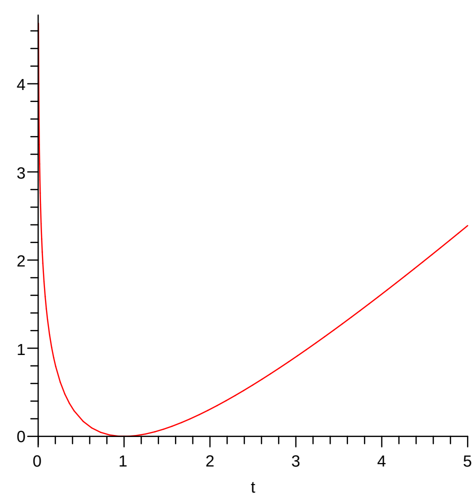

Figura 1.1: Gráfica de φ

Por la propiedad (a) y la definición de $d(x, y)$ se verifica que:

$$d(x, y) \geq 0 \text{ y } d(x, y) = 0 \Leftrightarrow x = y, \forall x, y \in \mathbb{R}^n_{++}.$$

Esto significa que d puede ser vista como una casi-distancia debido que se puede dar el caso donde d no cumpla la desigualdad triangular o la simetría. En realidad, d es una medida de divergencia en $\mathbb{R}^n_{++}$ (ver [5]) que cumple algunas propiedades, por ejemplo, las dadas en los dos lemas siguientes 1.2 y 1.3. En adición, sabemos que $d_\varphi : \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \longrightarrow \mathbb{R}$ es una función continua la cual puede ser continuamente extendida a $\mathbb{R}^n_{++} \times \mathbb{R}^n_{+}$ adoptando la convención de que $0 \ln 0 = 0$, es decir, d admite puntos con componente 0 en el segundo argumento.

1.3. Algunas propiedades de la función d

Lema 1.2. *Sea φ y d definidas como en (4) y (7), respectivamente. Entonces,*

(a) $d(x,z) - d(y,z) \geq \sum_{i=1}^{n} (z_i - y_i)\varphi^{'}(y_i/x_i)$ *para todo* $x, y \in \mathbb{R}^n_{++}$ *y* $z \in \mathbb{R}^n_{+}$.

(b) *Para un* $y \in \mathbb{R}^n_{+}$, *fijo ,* $L_x(y,\gamma) := \{x \in \mathbb{R}^n_{++} \mid d(x,y) \leq \gamma\}$ *son acotados para todo* $\gamma \geq 0$.

(c) *Para un* $x \in \mathbb{R}^n_{++}$, *fijo ,* $L_y(x,\gamma) := \{y \in \mathbb{R}^n_{+} \mid d(x,y) \leq \gamma\}$ *son acotados para todo* $\gamma \geq 0$.

Demostración. (a) Por definición de d podemos calcular

$$\begin{aligned} d(x,z) - d(y,z) &= \sum_{i=1}^{n} [z_i \ln(z_i/x_i) + x_i - z_i] - \sum_{i=1}^{n} [z_i \ln(z_i/y_i) + y_i - z_i] \\ &= \sum_{i=1}^{n} [z_i \ln(y_i/x_i) + x_i - y_i] \qquad (1.5) \end{aligned}$$

como $\varphi^{'}(y_i/x_i) = 1 - x_i/y_i$, tenemos que $y_i\varphi^{'}(y_i/x_i) = y_i - x_i$, es decir

$$x_i - y_i = -y_i\varphi^{'}(y_i/x_i). \qquad (1.6)$$

En adición, usando la propiedad (e) de φ y tomando en cuenta que $z_i \geq 0$ para todo $i = 1, \ldots, n$, obtenemos que

$$z_i \ln(y_i/x_i) \geq z_i\varphi^{'}(y_i/x_i). \qquad (1.7)$$

De las ecuaciones (1.5), (1.6) y (1.7), resulta inmediatamente que

$$d(x,z) - d(y,z) \geq \sum_{i=1}^{n} [z_i \varphi^{'}(y_i/x_i) - y_i \varphi^{'}(y_i/x_i)] = \sum_{i=1}^{n} (z_i - y_i)\varphi^{'}(y_i/x_i). \tag{1.8}$$

(b) Sea $\gamma \geq 0$ y $y \in \mathbb{R}^n_+$ fijo.
Supongamos que $y_i > 0$, para todo $i = 1, 2, \ldots, n$. Entonces

$$\begin{aligned} d(x,y) \leq \gamma \quad &\Rightarrow \quad y_i \varphi(x_i/y_i) \leq \gamma, \quad \forall i \in \{1, 2, \ldots, n\} \\ &\Rightarrow \quad \varphi(x_i/y_i) \leq \gamma/y_i, \quad \forall i \in \{1, 2, \ldots, n\}, \end{aligned}$$

así los conjuntos $L_x(y,\gamma) := \{x \in \mathbb{R}^n_{++} \mid d(x,y) \leq \gamma\}$ son acotados para todo $\gamma \geq 0$ con $y_i > 0$, para todo $i = 1, 2, \ldots, n$ por la propiedad (b).
Ahora supongamos que $y_i = 0$, para todo $i = 1, 2, \ldots, n$. Entonces

$$d(x,y) = d(x,0) = \sum_{i=1}^{n} 0 \ln(0/x_i) + x_i - 0 = \sum_{i=1}^{n} x_i,$$

luego, como $d(x,y) \leq \gamma$ entonces $\sum_{i=1}^{n} x_i \leq \gamma$, por lo que $x_i \leq \gamma$, para todo $i = 1, 2, \ldots, n$, de esto se tiene que $L_x(y,\gamma) := \{x \in \mathbb{R}^n_{++} \mid d(x,y) \leq \gamma\}$ son acotados para todo $\gamma \geq 0$ con $y_i = 0$, para todo $i = 1, 2, \ldots, n$.

(c) Sea $\psi(t) = t \ln t - t + 1$ con $t \geq 0$. Por definición de d tenemos que:
$d_\psi(x,y) = \sum_{i=1}^{n} x_i \ln(x_i/y_i) + y_i - x_i$, con $x \in \mathbb{R}^n_+$ y $y \in \mathbb{R}^n_{++}$. Además

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

Así,

$$d_\psi(x,y) = d_\varphi(y,x). \tag{1.9}$$

Sea $x \in \mathbb{R}^n_{++}$ fijo. Probar que $L_y(x,\gamma) := \{y \in \mathbb{R}^n_+ \mid d_\varphi(x,y) \leq \gamma\}$ son acotados para todo $\gamma \geq 0$ es equivalente a probar que $L_y(x,\gamma) := \{y \in \mathbb{R}^n_+ \mid d_\psi(y,x) \leq \gamma\}$ son acotados para todo $\gamma \geq 0$ por la igualdad en (1.9), lo cual fue probado en la parte (b) ya que $\lim_{k\to+\infty} \psi(t) = +\infty$. $\square$

Lema 1.3. *Dadas dos sucesiones cualesquiera $\{y^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^n_{++}$ y $\{x^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^n_+$,*

(a) *si $\{y^k\}_{k\in\mathbb{N}}$ converge a $\bar{y} \in \mathbb{R}^n_+$, entonces* $\lim_{k\to+\infty} d(y^k,\bar{y}) = 0$.

(b) *si $\{y^k\}_{k\in\mathbb{N}}$ es acotada y $\{x^k\}_{k\in\mathbb{N}}$ es tal que* $\lim_{k\to+\infty} d(y^k,x^k) = 0$ *, entonces tenemos que* $\lim_{k\to+\infty} \|y^k - x^k\| = 0$.

Demostración. (a) De la definición de d, se sigue que

$$d(y^k,\bar{y}) = \sum_{i=1}^{n} [\bar{y}_i \ln \bar{y}_i - \bar{y}_i \ln y_i^k + (y_i^k - \bar{y}_i)]$$

Supongamos que $\bar{y}_i = 0$, claramente

$$\bar{y}_i \ln \bar{y}_i - \bar{y}_i \ln y_i^k + (y_i^k - \bar{y}_i) = 0 \ln 0 - 0 \ln y_i^k + (y_i^k - 0) = y_i^k.$$

Así,

$$\lim_{k\to+\infty} [\bar{y}_i \ln \bar{y}_i - \bar{y}_i \ln y_i^k + (y_i^k - \bar{y}_i)] = \lim_{k\to+\infty} y_i^k = \bar{y}_i = 0.$$

Ahora supongamos que $\bar{y}_i > 0$, entonces

$$\lim_{k\to+\infty} \ln(\bar{y}_i/y_i^k) = \ln(\lim_{k\to+\infty} \bar{y}_i / \lim_{k\to+\infty} y_i^k) = \ln(\bar{y}_i/\bar{y}_i) = \ln(1) = 0$$

y $\lim_{k\to+\infty} (y_i^k - \bar{y}_i) = \bar{y}_i - \bar{y}_i = 0$. Así,

$$\lim_{k\to+\infty} [\bar{y}_i \ln \bar{y}_i - \bar{y}_i \ln y_i^k + (y_i^k - \bar{y}_i)] = \lim_{k\to+\infty} [\bar{y}_i \ln(\bar{y}_i/y_{i^k}) + (y_i^k - \bar{y}_i)] = 0.$$

Por ambas partes concluimos que

$$\lim_{k\to+\infty} d(y^k,\bar{y}) = 0.$$

(b)En primer lugar probemos que la sucesión $\{x^k\}_{k\in\mathbb{N}}$ es acotada, supongamos lo contrario, que no es acotada. Ahora bien sabemos por hipótesis que

$$\lim_{k\to+\infty} d(y^k,x^k) = \lim_{k\to+\infty} \sum_{i=1}^{n} x_i^k \varphi(y_i^k/x_i^k) = 0,$$

es decir la sucesión $\{x_i^k\varphi(y_i^k/x_i^k)\}_{k\in\mathbb{N}}$ es convergente para todo $i \in \{1,2,..,n\}$ y por tanto acotada, entonces existe un $M > 0$ tal que

$$\begin{aligned} x_i^k\varphi(y_i^k/x_i^k) &\leq M \qquad \forall i \in \{1,2,..,n\} \\ 0 \leq \varphi(y_i^k/x_i^k) &\leq M/x_i^k \qquad \forall i \in \{1,2,..,n\} \end{aligned} \tag{1.10}$$

Ahora bien, $y_i^k/x_i^k \to 0$ cuando $k \to +\infty$ ya que $\{y_i^k\}_{k\in\mathbb{N}}$ es acotada $\forall i \in \{1,2,\ldots,n\}$. Así cuando $k \to +\infty$ la ecuación (1.10) nos queda

$$0 \leq \varphi(0) \leq 0,$$

lo cual contradice la definición de φ . Por tanto la sucesión $\{x^k\}_{k\in\mathbb{N}}$ es acotada.

Ahora por el método de reducción al absurdo, supongamos que $\lim_{k\to+\infty} \|y^k - x^k\| \neq 0$. Entonces existen dos subsucesiones $\{y^{\sigma(k)}\}_{k\in\mathbb{N}}$ y $\{x^{\sigma(k)}\}_{k\in\mathbb{N}}$ de $\{y^k\}_{k\in\mathbb{N}}$ y $\{x^k\}_{k\in\mathbb{N}}$ respectivamente tales que $\|y^{\sigma(k)} - x^{\sigma(k)}\| \geq 3\varepsilon$ para $\varepsilon > 0$ y k suficientemente grande. Ahora bien, como $\{y^k\}_{k\in\mathbb{N}}$ y $\{x^k\}_{k\in\mathbb{N}}$ son acotadas entonces las subsucesiones $\{y^{\sigma(k)}\}$ y $\{x^{\sigma(k)}\}$ son convergentes, digamos sin perdida de generalidad que $\{y^{\sigma(k)}\}_{k\in\mathbb{N}} \to y^*$ y $\{x^{\sigma(k)}\}_{k\in\mathbb{N}} \to x^*$. De la desigualdad triangular tenemos lo siguiente

$$\|y^{\sigma(k)} - y^*\| + \|y^* - x^{\sigma(k)}\| \geq \|y^{\sigma(k)} - x^{\sigma(k)}\| \geq 3\varepsilon. \tag{1.11}$$

Como $\{y^{\sigma(k)}\}_{k\in\mathbb{N}} \to y^*$, entonces existe un entero positivo K tal que para $k \geq K$, $\|y^{\sigma(k)} - y^*\| \leq \varepsilon$. Así de (1.11) tenemos que $\|y^* - x^{\sigma(k)}\| \geq 3\varepsilon - \|y^{\sigma(k)} - y^*\| \geq 2\varepsilon$ para $k \geq K$. Por otro parte, como $\{x^{\sigma(k)}\}_{k\in\mathbb{N}} \to x^*$ para $k \geq K$, $\|x^{\sigma(k)} - x^*\| \leq \varepsilon$, luego

$$\begin{aligned} 2\varepsilon \leq \|y^* - x^{\sigma(k)}\| &= \|y^* - x^* + x^* - x^{\sigma(k)}\| \\ &\leq \|y^* - x^*\| + \|x^{\sigma(k)} - x^*\| \\ &\leq \|y^* - x^*\| + \varepsilon. \end{aligned}$$

Por lo que

$$\|y^* - x^*\| \geq 2\varepsilon - \varepsilon = \varepsilon. \tag{1.12}$$

Pero, como $\lim_{k\to+\infty} d(y^k, x^k) = 0$, ocurre que $\lim_{k\to+\infty} d(y^{\sigma(k)}, x^{\sigma(k)}) = 0$ por tanto por la continuidad de d se sigue que $d(y^*, x^*) = 0$, de modo que $y^* = x^*$ lo cual es una contradicción con 1.12. $\square$

Capítulo 2

Análisis de Convergencia

Es esta sección, estableceremos los resultados de convergencia del algoritmo casi-proximal (6)-(7). Primero, mostraremos que el algoritmo esta bien definido bajo la siguiente suposición:
(A1) El conjunto óptimo del problema (5), denotado por $\mathcal{X}^*$, es no vacío y acotado.

Lema 2.1. *Sea d definida como en (7). Entonces bajo la suposición (A1),*

(a) *La sucesión $\{x^k\}_{k\in\mathbb{N}}$ generada por (6)-(7) esta bien definida.*

(b) *$\{f(x^k)\}_{k\in\mathbb{N}}$ es decreciente y convergente.*

Demostración. (a) La prueba se hará por inducción. Claramente, cuando $k = 0$, la conclusión se cumple puesto que $x^0 > 0$ es dado. Supongamos que x^{k-1} esta bien definida. Sea f^* el valor óptimo de (5). Entonces, del esquema iterativo (6) y no negatividad de d, se sigue que

$$f(x) + \mu_k d(x, x^{k-1}) \geq f^* + \mu_k d(x, x^{k-1}) \qquad \forall x \in \mathbb{R}^n_{++}. \tag{2.1}$$

Sea $f_k(x) := f(x) + \mu_k d(x, x^{k-1})$ y denotemos sus conjuntos de nivel por

$$L_{f_k}(\gamma) := \{x \in \mathbb{R}^n_{++} : f_k(x) \leq \gamma\} \qquad \text{para algún} \quad \gamma \in \mathbb{R}. \tag{2.2}$$

Usando la desigualdad en (2.1), tenemos $L_{f_k}(\gamma) \subseteq L_x(x^{k-1}, \mu_k^{-1}(\gamma - f^*))$, en efecto:

$$\begin{aligned}
x \in L_{f_k}(\gamma) \;&\Rightarrow\; f_k(x) \leq \gamma, \;\; x \in \mathbb{R}^n_{++} \\
&\Rightarrow\; f(x) + \mu_k d(x, x^{k-1}) \leq \gamma, \;\; x \in \mathbb{R}^n_{++} \\
&\Rightarrow\; d(x, x^{k-1}) \leq (\gamma - f(x))\mu_k^{-1}, \;\; x \in \mathbb{R}^n_{++} \\
&\Rightarrow\; d(x, x^{k-1}) \leq (\gamma - f^*)\mu_k^{-1} \;\; x \in \mathbb{R}^n_{++} \\
&\Rightarrow\; x \in L_x(x^{k-1}, \mu_k^{-1}(\gamma - f^*)).
\end{aligned}$$

Esto junto con el Lema 1.2 (b), implica que $L_{f_k}(\gamma)$ es acotado para $\gamma \geq f^*$. Notemos que si $\gamma \leq f^*$, entonces $L_{f_k}(\gamma) \subset X^*$, en efecto:

$$\begin{aligned}
x \in L_{f_k}(\gamma) \;&\Rightarrow\; f(x) + \mu_k d(x, x^{k-1}) \leq \gamma, \;\; x \in \mathbb{R}^n_{++} \\
&\Rightarrow\; f(x) + \mu_k d(x, x^{k-1}) \leq \gamma \leq f^*, \;\; x \in \mathbb{R}^n_{++} \\
&\Rightarrow\; f(x) \leq f(x) + \mu_k d(x, x^{k-1}) \leq \gamma \leq f^*, \;\; x \in \mathbb{R}^n_{++} \\
&\Rightarrow\; f(x) \leq f^*, \;\; x \in \mathbb{R}^n_{++} \\
&\Rightarrow\; f(x) = f^*, \;\; x \in \mathbb{R}^n_{++} \\
&\Rightarrow\; x \in \mathcal{X}^*.
\end{aligned}$$

En consecuencia $L_{f_k}(\gamma)$ también es acotado para $\gamma \leq f^*$ por la suposición (A1). Esto demuestra que los conjuntos de nivel de $f_k(x)$ son acotados, además como $f_k(x)$ es semi-continua inferiormente en $\mathbb{R}^n$, tenemos que los conjuntos de nivel de $f_k(x)$ son compactos. Usando nuevamente la semi-continuidad inferior de $f_k(x)$, tenemos que $f_k(x)$ tiene un mínimo global el cual no necesariamente es único por la no-convexidad de f. En tal caso , x^k puede ser elegida arbitrariamente entre el conjunto de minimizadores de $f_k(x)$. Así la sucesión $\{x^k\}_{k\in\mathbb{N}}$ está bien definida.

(b) Del esquema iterativo en (6)

$$f(x^k) + \mu_k d(x^k, x^{k-1}) \leq f(x) + \mu_k d(x, x^{k-1}), \;\; \forall x \in \mathbb{R}_{++}, \tag{2.3}$$

haciendo $x = x^{k+1}$ en la anterior desigualdad, obtenemos que

$$f(x^k) + \mu_k d(x^k, x^{k-1}) \leq f(x^{k-1}) + \mu_k d(x^{k-1}, x^{k-1}) = f(x^{k-1}),$$

lo cual por la no-negatividad de d y μ_k, implica que

$$0 \leq \mu_k d(x^k, x^{k-1}) \leq f(x^{k-1}) - f(x^k).$$

Así,

$$f(x^k) \leq f(x^{k-1}).$$

Esto muestra que $\{f(x^k)\}_{k\in\mathbb{N}}$ es una sucesión decreciente y como es acotada por debajo es convergente. □

Por el lema 2.1 (b), sea $\beta = \lim_{k\to+\infty} f(x^k) = \inf_{x^k\in\mathbb{R}^n_+} f(x^k)$ y definamos el siguiente conjunto:

$$\mathcal{U} := \{x \in \mathbb{R}^n_+ | f(x) \leq \beta\}. \tag{2.4}$$

Notemos que $\mathcal{X}^* \subseteq \mathcal{U}$, en efecto,

Sea $A = \{x^k : x^k \textit{ es generado por el algoritmo } (6) \textit{ y } (7)\}$, por definición $A \subseteq \mathbb{R}^n_+$, así que

$$f(A) = \{f(x^k) : x^k \textit{ es generado por el algoritmo } (6) \textit{ y } (7)\} \subseteq \{f(x) : x \in \mathbb{R}^n_+\}$$

por tanto

$$\inf_{x^k\in\mathbb{R}^n_+} (A) \geq \inf_{x\in\mathbb{R}^n_+} \{f(x) : x \in \mathbb{R}^n_+\}. \tag{2.5}$$

Ahora bien,

$$\begin{aligned} x \in \mathcal{X}^* \quad &\Rightarrow \quad f(x) = \inf_{x\in\mathbb{R}^n_+} f(x) \\ &\Rightarrow \quad f(x) \leq \inf_{x^k\in\mathbb{R}^n_+} f(x^k) = \beta \\ &\Rightarrow \quad x \in \mathcal{U}. \end{aligned}$$

Por lo que $\mathcal{X}^* \subseteq \mathcal{U}$ y en consecuencia $\mathcal{U}$ es no vacío por la suposición (A1). En lo que sigue mostraremos que la sucesión generada por (6) es Fejér convergente en $\mathcal{U}$ con respecto a d bajo la siguiente suposición adicional:

(A2) f es una función casi-convexa diferenciable en el $dom f$.

Lema 2.2. *Sea $\{\mu_k\}_{k\in\mathbb{N}}$ una sucesión arbitraria de números positivos y $\{x^k\}_{k\in\mathbb{N}}$ la sucesión generada por (6)-(7). Entonces bajo la suposición (A1) y (A2),*

(a) $d(x^k, x) \leq d(x^{k-1}, x), \quad \forall x \in \mathbb{R}^n_+$ *que cumpla que* $f(x) \leq f(x^k)$.

(b) $\{x^k\}_{k\in\mathbb{N}}$ *es Fejér convergente al conjunto U con respecto a d.*

(c) *Para todo $x \in U$, la sucesión $\{d(x^k, x)\}_{k\in\mathbb{N}}$ es convergente.*

Demostración. (a) Sea $x \in \mathbb{R}^n_+$ que satisface que $f(x) \leq f(x^k)$; del Lema 1.1 (b) se sigue que

$$\nabla f(x^k)^T (x - x^k) \leq 0. \tag{2.6}$$

En adición, como x^k es minimizador de la función $f(x) + \mu_k d(x, x^{k-1})$, tenemos

$$\nabla f(x^k) + \mu_k \nabla_x d(x, x^{k-1}) = 0. \tag{2.7}$$

Combinando las ecuaciones (2.6) y (2.7) , obtenemos que

$$\mu_k \langle \nabla_x d(x^k, x^{k-1}), x - x^k \rangle \geq 0 \qquad \forall x \in \mathbb{R}^n_+ \text{ que satisface } f(x) \leq f(x^k).$$

Notemos que

$$\nabla_x d(x, x^{k-1}) = (\varphi^{'}(x_1^k/x_1^{k-1}), ..., \varphi^{'}(x_n^k/x_n^{k-1}))^T,$$

luego usando el lema 1.2 (a) con $x = x^{k-1}$, $y = x^k$ y $z = x$, se sigue que

$$d(x^{k-1}, x) - d(x^k, x) \geq \mu_k \langle \nabla_x d(x^k, x^{k-1}), x - x^k \rangle \geq 0,$$

así, $d(x^k, x) \leq d(x^{k-1}, x)$, para un $x \in \mathbb{R}^n_+$ que satisface $f(x) \leq f(x^k)$.

(b) Por definición de U tenemos que si $x \in U$ entonces $f(x) \leq f(x^k)$ para todo k, en efecto

$$x \in U \Rightarrow f(x) \leq \beta = \lim_{k \to +\infty} f(x^k), \quad x \in \mathbb{R}^n_+ \tag{2.8}$$

pero como $\{f(x^k)\}_{k \in \mathbb{N}}$ es decreciente se tiene que $\lim_{k \to +\infty} f(x^k) \leq f(x^k)$, luego de (2.8) obtenemos que

$$x \in U \Rightarrow f(x) \leq f(x^k), \quad x \in \mathbb{R}^n_+.$$

Usando la parte (a) tenemos que $d(x^k, x) \leq d(x^{k-1}, x)$ con $x \in U$, así por la definición (1.2) se concluye que d es Fejér convergente en U.

(c) De la parte (a) tenemos que $d(x^k, x) \leq d(x^{k-1}, x)$ así $\{d(x^k, x)\}_{k \in \mathbb{N}}$ es decreciente, ahora bien como d es no-negativa ocurre que $0 \leq d(x^k, x) \leq d(x^{k-1}, x)$, luego $\{d(x^k, x)\}_{k \in \mathbb{N}}$ es decreciente y acotada por debajo, por tanto $\{d(x^k, x)\}_{k \in \mathbb{N}}$ es convergente. $\square$

2.1. Teorema Central

Teorema 2.1. *Sea $\{\mu_k\}_{k\in\mathbb{N}}$ una sucesión arbitraria de números positivos y $\{x^k\}_{k\in\mathbb{N}}$ generada por (6)-(7). Si las suposiciones (A1) y (A2) se cumplen, entonces $\{x^k\}_{k\in\mathbb{N}}$ converge y*

(a) *Si existe $\hat{\mu}$ y $\bar{\mu}$ tales que $o < \hat{\mu} < \mu_k < \bar{\mu}$ para cada $k \in \mathbb{N}$, entonces $\lim\limits_{k\to+\infty}(\nabla f(x^k))_i \geq 0$, $\lim\limits_{k\to+\infty}(\nabla f(x^k))_i(x_i - x_i^k) \geq 0$ para todo $x \in \mathbb{R}^n_+$, $i \in \{1, 2, ..., n\}$.*

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

Demostración. Primero probemos que la sucesión $\{x^k\}_{k\in\mathbb{N}}$ es convergente. Por el Lema 2.2 (b), $\{x^k\}_{k\in\mathbb{N}}$ es Fejér convergente en el conjunto $\mathcal{U}$ con respecto a d, lo cual implica que

$$\{x^k\}_{k\in\mathbb{N}} \subseteq \{y \in \mathbb{R}^n_{++} | d(y, x) \leq d(x^0, x)\}, \quad \forall x \in \mathcal{U}.$$

Del Lema 1.2 (b), tenemos que el conjunto del lado derecho de la ecuación anterior es acotado y en consecuencia $\{x^k\}_{k\in\mathbb{N}}$ es acotada. Sea $\bar{x}$ un punto de acumulación de $\{x^k\}_{k\in\mathbb{N}}$ y $\{x^{k_j}\}_{k\in\mathbb{N}}$ una subsucesión de $\{x^k\}_{k\in\mathbb{N}}$ que converge a $\bar{x}$. De la continuidad de f, se sigue que

$$\lim_{j\to\infty} f(x^{k_j}) = f(\bar{x}),$$

el cual por definición de $\mathcal{U}$, implica que $\bar{x} \in \mathcal{U}$. Usando el Lema 2.2 (c), tenemos que la sucesión $\{d(x^k, \bar{x})\}_{k\in\mathbb{N}}$ es convergente. Notemos que $\lim\limits_{j\to+\infty}\{d(x^{k_j}, \bar{x})\} = 0$ por Lema 1.3 (a), por tanto $\lim\limits_{k\to+\infty}\{d(x^k, \bar{x})\} = 0$. Usando el Lema 1.3 (b) con $y^k = x^k$ y $x^k = \bar{x}$, tenemos que $\lim\limits_{t\to+\infty} \|x^k - \bar{x}\| = 0$, de esto $\{x^k\}_{k\in\mathbb{N}}$ converge a $\bar{x}$.

(a) Por la fórmula iterativa en (6), x^k es minimizador de $f(x) + \mu_k d(x, x^{k-1})$, por lo que $\nabla f(x^k) + \nabla_x d(x^k, x^{k-1}) = 0$, de esto obtenemos

$$\nabla f(x^k) = -\nabla_x d(x^k, x^{k-1}),$$

esto significa que

$$(\nabla f(x^k))_i = -\mu_k \varphi'(x_i^k / x_i^{k-1}), \quad i \in \{1, 2, ..., n\}. \tag{2.9}$$

Sea $\bar{x}$ el limite de la sucesión $\{x^k\}_{k\in\mathbb{N}}$. Definamos los conjuntos de indices

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

Claramente, los dos conjuntos son disjuntos y forman una partición del conjunto $\{1,2,...,n\}$. Analicemos los casos donde $i \in I(\bar{x})$ o $i \in J(\bar{x})$ para $i \in \{1,2,...,n\}$.

Caso(1): Si $i \in I(\bar{x})$, entonces $\lim\limits_{k\to+\infty}(x_i^k/x_i^{k-1}) = 1$ por la convergencia de $\{x^k\}_{k\in\mathbb{N}}$. Por la continuidad de φ' en $(0,+\infty)$ y como $\varphi'(1) = 0$, tenemos que

$$\lim_{k\to+\infty} \varphi'(x_i^k/x_i^{k-1}) = 0,$$

esto junto con (2.9) y como $\{\mu_k\}_{k\in\mathbb{N}}$ es acotada, obtenemos inmediatamente que

$$\lim_{k\to+\infty} (\nabla f(x^k))_i = 0$$

en consecuencia por la convergencia de $\{x^k\}_{k\in\mathbb{N}}$, se tiene

$$\lim_{k\to+\infty} (\nabla f(x^k))_i (x_i - x_i^k) = 0$$

Caso(2): Sea $i \in J(\bar{x})$. Supongamos por reducción al absurdo que $\lim\limits_{k\to+\infty}(\nabla f(x^k))_i < 0$ para $k \in \mathbb{N}$ suficientemente grande . De (2.9) y como $\{\mu_k\}_{k\in\mathbb{N}}$ es una sucesión de números positivos se tiene $\varphi'(x_i^k/x_i^{k-1}) > 0$ para $k \in \mathbb{N}$ suficientemente grande, luego por las propiedades (c) y (d) de φ, $x_i^k > x_i^{k-1}$ para $k \in \mathbb{N}$ suficientemente grande, lo cual contradice que

$$x^k \to \bar{x}.$$

por consiguiente, $\lim\limits_{k\to+\infty}(\nabla f(x^k))_i \geq 0$. Notemos que $\lim\limits_{k\to+\infty}(x_i - x_i^k) \geq 0$ como $x_i \geq 0$ y $\lim\limits_{k\to+\infty} x_i^k = \bar{x}_i = 0$, así deducimos que $\lim\limits_{k\to+\infty}(\nabla f(x^k))_i(x_i - x_i^k) \geq 0$.

(b) Como x^k es minimizador de $f(x) + \mu_k d(x, x^{k-1})$, tenemos

$$f(x^k) + \mu_k d(x^k, x^{k-1}) \leq f(x) + \mu_k d(x, x^{k-1}), \qquad \forall x \in \mathbb{R}^n_{++},$$

lo cual por la no negatividad de d y como $\{\mu_k\}_{k\in\mathbb{N}}$ es una sucesión de números positivos, obtenemos

$$f(x^k) \leq f(x) + \mu_k d(x, x^{k-1}), \qquad \forall x \in \mathbb{R}^n_{++},$$

tomando el limite cuando $k \to +\infty$ en la desigualdad anterior y usando la continuidad de f ocurre que

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

ya que $\lim_{k\to+\infty} x^k = \bar{x}$, $\lim_{k\to+\infty} \mu_k = 0$ y la sucesión $\{d(x, x^{k-1})\}_{k\in\mathbb{N}}$ es acotada por el Lema 1.2 (c). Esto significa que $f(\bar{x}) \leq f(x)$ para todo $x \in \mathbb{R}^n_+$, en consecuencia, $\bar{x} \in \mathcal{X}^*$. Así la prueba esta completa.

$\square$

Conclusión

En este trabajo consideramos el método casi proximal definido en (6)-(7) para resolver un problema no-convexo de la forma (5) y establecimos los resultados de convergencia del algoritmo bajo algunas suposiciones. Específicamente, bajo la suposición de que el conjunto óptimo del problema (5), denotado por $\mathcal{X}^*$, es no vacío y acotado y f es una función casi-convexa diferenciable en el $dom f$, la sucesión $\{x^k\}_{k\in\mathbb{R}}$ generada por el algoritmo es convergente y además converge a una solución óptima del problema (5).

REFERENCIAS

[1] M. S. BAZARAA AND C. M. SHETTY, *Nolinear Programming Theory and Algomithms*, New York: John Wiley and Sons, 1979.

[2] S. CHRETIEN AND O. HERO, *Generalized proximal point algorithms, Technical Report*, The University Of Michigan, 1998.

[3] G. CHEN AND M. TEBOULLE,*Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM Jornual on Optimization, vol. 3 (1993), pp. 538-543.

[4] P. B. EGGERMONT, *Multilicative iterative algorithms for convex programming*, Linear Algebra Appl. vol. 130 (1990), pp. 403-419

[5] A. IUSEM AND M. TEBOULLE. *Convergence rate analysis of nonquadratic proximal method for convex and linear programming*, Mathematics and Operation Research. vol. 20 (1995) pp. 657-677.

[6] R. T. ROCKAFELLAR, *Augmented Lagrangians and applications of proximal point algorithm convex programming*, Mathematics of Operation Research, vol. 1 (1976), pp. 97-116.

[7] R. T. ROCKAFELLAR, *Monotone Operators and the Proximal Point algorithm*, SIAM Jornual on Control and Optimization, vol. 14 (1976), pp. 877-898.

[8] A. IUSEM, *Métodos de Ponto Proximal em optimizaão* , 20 Colóquio Brasileiro de matemática. IMPA. 24-28 Julho, (1995)

[9] M. TEBOULLE , *Entropic proximal mappings with applications to nonlinear programming* , Mathematics of Operation Research, vol. 17 (1192), pp. 670-690 97-116.