UNIVERSIDAD CENTROCCIDENTAL
"LISANDRO ALVARADO"

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas

"UN ALGORITMO DE PUNTO PROXIMAL CON $\varphi-$DIVERGENCIA PARA PROGRAMACIÓN CUASICONVEXA"

Trabajo Especial de Grado presentado por

Br. María L. Cortéz M.

como requisito final
para obtener el título de Licenciado
en Ciencias Matemáticas
Área de Conocimiento: Optimización.
Tutor: MCs. Jorge Campos.

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:

"UN ALGORITMO DE PUNTO PROXIMAL CON $\varphi-$DIVERGENCIA PARA PROGRAMACIÓN CUASICONVEXA"

Presentado por la ciudadana Br. María L. Cortéz M. titular de la Cédula de Identidad Nº 17.306.111. 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 Dios Todopoderoso y a mis
Padres: Isidro y Yolanda.

AGRADECIMIENTOS

Primeramente doy **Gracias a Dios**, Todopoderoso, por ser mi guía, mi proveedor durante toda mi vida, aquel que me permite abrir los ojos y respirar día a día. Padre celestial tu me inspiras, y me das la fortaleza para seguir y luchar por mis sueños y ser el Matemático por excelencia, sin duda, sin ti no habría llegado hasta este momento. Todo lo puedo en Cristo que me fortalece.

A **mi madre**, serás siempre mi inspiración para lograr mis objetivos. Gracias mami por todo tu amor incondicional e infinito, y por siempre estar cuando más te necesito. Mi orgullo eres tu mami, lo que hoy por hoy soy, te lo debo a ti y por ello te dedico este logro.¡Gracias Mamá!

A **mi padre**, por todo el amor y apoyo que me ha brindado. Excelente padre. Gracias por estar presente en todos los momentos de mi vida.¡Gracias Papá!

A mis hermanos, **Nohíri**, **Yelitza**, **Lisbet**, **César** y **Yolier**, gracias por siempre estar ahí presentes en todo momento, por siempre confiar en mí y nunca dudar que lo lograría. Gracias por todo lo que han hecho por mi.

A mi abuela **María de los Reyes**, quién de manera silenciosa siempre me ha apoyado.

A **mis pedacitos de cielo**, mis sobrinos, Arnaldo, Isamar y Móises, quienes con sus juegos y su amor logran apartar de mí las tristezas y preocupaciones. ¡Los Amo mucho mis niños!

A **Juan José Gónzalez**, quién siempre me brindo su comprensión, estímulo y apoyo constante durante toda mi carrera. Gracias por tu paciencia, y por enseñarme tantas cosas importantes en mi vida. Este logro también es tuyo.

A mi gran amiga **Orlaidé**, que ha sido una hermana y me ha acompañado durante toda mi carrera, compartiendo mis mejores momentos, y lloró conmigo en mis tragos amargos. Gracias amiga por haber estado a mi lado, eres una bendición y estoy muy orgulloso de ser tu amiga.

A **Carlos Luis**, quien creyó en mi y me apoyo a finales de mi carrera. Gracias por tu gran ayuda eres una persona muy especial para mi, sin ti todo fuese sido más difícil.

A mis amigas y compañeras,**Gabriela Galavis**, **Gabriela Gónzalez**, que han sido también como unas hermanas para mí. Gracias por su amistad sincera e incondicional, y por todas las cosas bellas que hemos vivido juntas. A pesar de la distancia y de los caminos que cada una tomo, siempre las

tengo presentes, y para mi es un privilegio llegar al final de la meta junto a ustedes y compartir esta gran felicidad.

A mis amigos, **Freddy y José**, por sus consejos, por estar siempre ahí dispuestos a ayudar, amigos ¡ánimo!, pronto serán ustedes quienes escribirán lineas cómo estas.

A **Juan Brizuela**, quien con su constancia y disciplina termino de formar mis hábitos de estudio. Gracias Juancitó por todo tu apoyo y tus consejos. Le doy gracias a Dios por compartir esta felicidad juntos.

A **María Teresa**, gracias por apoyarme y por darme aliento cuando más lo necesité. ¡Te quiero mucho Tere!

A **Johela y Yorisbel**, gracias amigas por estar ahí siempre pendientes de mi y por brindarme una amistad genuina e incondicional. Sus ayudas fueron muy importantes para mi, las quiero mucho.

A mi tutor, **Jorge Campos**, por su asesoría y dirección en el cumplimiento de este proyecto, gracias por su disposición y su paciencia ante mi inexperiencia. Gracias por su confianza, sinceridad y palabras de aliento, gracias por todo este tiempo que me dedicó. Estaré en deuda con usted toda mi vida.

A los **profesores** que contribuyeron en mi formación profesional, enseñándome todas las cosas que necesitaba y por sus grandes consejos y ayuda. A los Profesores: Eibar, Rómulo Castillo, Mario, Javier, Omar, Ebner, Miguel, Luz Rodríguez, Alexander Carrasco. A todos, ¡Gracias por compartir sus conocimientos!

A la **UCLA** por haberme formado académicamente.

Y en especial les doy las gracias a dos personas que siempre me ayudaron y me apoyaron durante mi carrera, y podría decir que sin ellas me habría costado mucho más. ¡Gracias **José Antonio Navas**!¡Gracias **Elvis**! Por todo lo que hicieron por mí, que Dios las bendiga en todas sus metas en la vida.

Y a todos los que de una manera u otra contribuyeron a este logro y ser la persona quien hoy soy, no tengo palabras para decirles lo que siento, solo les puedo decir: **¡Muchísimas Gracias!**

Resumen

En este trabajo se pretende estudiar y desarrollar el método de punto proximal con $\varphi-$divergencia dada por $\varphi(t) = t - \log t - 1$ para la minimización de funciones cuasiconvexas sujetas a restricciones de no negatividad que fueron publicadas en un artículo por Cunha, Da Cruz Neto y Oliveira en junio del 2006, estableciendo que la sucesión generada por el algoritmo está bien definida en el sentido de que existe y no es cíclica. Sin ningún tipo de hipótesis de nivel de acotación de la función objetivo, se obtiene que la sucesión converge a un punto estacionario. También demostraremos que cuando los parámetros de regularización van a cero, la sucesión converge a una solución óptima.

Introducción

Consideremos el problema de optimización cuasiconvexa

$$(P) \qquad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & x \geq 0, \end{array}$$

donde $f : \mathbb{R}^n \to (-\infty, +\infty]$ es una función cuasiconvexa apropiada (es decir, $dom f := \{x \in \mathbb{R}^n : f(x) < +\infty\} \neq \emptyset$).

La clase interesante de funciones cuasiconvexas (o eventualmente sus subclases como las clases de funciones estríctamente o fuertemente cuasiconvexas) generaliza funciones convexas manteniendo algunas de sus propiedades más importantes. Estas funciones disfrutan de notables propiedades de estabilidad y tiene un gran dominio en las aplicaciones de diversos campos de las ciencias y la ingeniería tales como la teoría económica, teoría de la localización, la teoría del control y la teoría de aproximación.

El algoritmo clásico de punto proximal, para minimizar una función convexa f en $\mathbb{R}^n$, genera una sucesión $\{x^k\}$ por el esquema iterativo: comenzar con un punto inicial $x^0 \in \mathbb{R}^n$ y resolver

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

donde $\{\lambda_k\}$ es una sucesión de números positivos. Este método fue introducido originalmente por Martinet [7] y desarrollado y estudiado por Rockafellar [8, 9]. La convergencia de la sucesión generada $\{x^k\}$ ha atraído la atención de muchos autores para el caso convexo.

Algunos investigadores han considerado la posibilidad de sustituir el término cuadrático usual en (1) por otros tipos de medidas tales como las distancias de Bregman o

por $\varphi-$divergencia.
El algoritmo de punto proximal con $\varphi-$divergencia fue introducido en 1992 por Teboulle [10]. En ella, el término cuadrático de (1) es sustituido por la distancia entropic-like

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

con φ satisfaciendo algunas condiciones, que presentaremos en la sección 2.2. Por lo tanto, el algoritmo de punto proximal con $\varphi-$divergencia para minimizar una función convexa f en el ortante no negativo de $\mathbb{R}^n$ viene dado por:

$$\begin{aligned} x^0 &> 0 \\ x^{k+1} &= \arg\min_{x\geq 0}\{f(x) + \lambda_k d_\varphi(x,x^k)\}, \quad \lambda_k > 0. \end{aligned} \tag{2}$$

El algoritmo (2) se ha estudiado ampliamente para la programación convexa, ver [4, 6, 10, 11] y sus referencias. Es importante señalar que la principal ventaja del algoritmo (2) en relación a el algoritmo (1) es que el término d_φ se utiliza para forzar a las iteraciones x^k a permanecer en el interior del ortante no negativo de $\mathbb{R}^n$, es decir, el algoritmo (2) va a generar automáticamente una sucesión positiva $\{x^k\}$.
En este trabajo se aplicará el algoritmo de punto proximal (2) con la $\varphi-$divergencia dada por $\varphi(t) = t - \log t - 1$ para resolver el problema (P).
Nuestro objetivo principal consiste en estudiar que la sucesión $\{x^k\}$ generada por el algoritmo esté bien definida y converge a un punto estacionario cuando el parámetro λ_k satisface

$$0 < \lambda_k \leq \tilde{\lambda}, \tag{3}$$

para algún $\tilde{\lambda} > 0$ (lo que se incluye el caso de λ_k constante). Asimismo, con el parámetro λ_k satisfaciendo la condición de regularidad

$$\lim_{k\to+\infty} \lambda_k = 0, \tag{4}$$

obtenemos entonces la convergencia a una solución del problema (P).

Índice

Capítulo 1

Preliminares.

El objetivo de este capítulo es dar una breve presentación de algunos conceptos y resultados auxiliares que nos permiten establecer las condiciones básicas para una buena comprensión del trabajo y lograr un trabajo autocontenido.

§1.1. Convexidad.

Definición 1.1. *Un conjunto* $S \subset \mathbb{R}^n$ *es llamado* ***convexo*** *si*

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

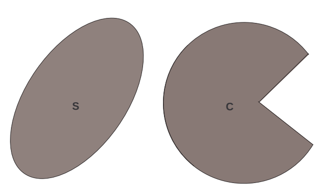

Figura 1.1: S es un conjunto convexo pero C no es un conjunto convexo.

Definición 1.2. *Sea* f *una función a valores reales extendidos con dominio* $S \subset \mathbb{R}^n$. *El conjunto*

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

es llamado el ***epígrafo*** *de* f.

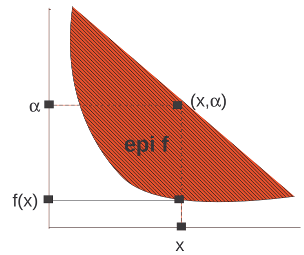

Figura 1.2: epígrafo.

Definición 1.3. *Sea $S \subset \mathbb{R}^n$ un conjunto convexo. Una función $f : S \to \mathbb{R}$ es* ***convexa*** *sobre S si*

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

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

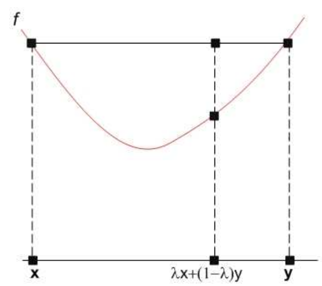

Figura 1.3: Función convexa.

Definición 1.4. *Sea $S \subset \mathbb{R}^n$ un conjunto convexo. Una función $f : S \to \mathbb{R}$ es* ***estríctamente convexa*** *sobre S si*

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y),$$

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

Definición 1.5. *Una función convexa f es llamada* ***propia*** *si su epígrafo es no vacío y no contiene rectas verticales, es decir, si $f(x) < +\infty$ para al menos un valor de x y $f(x) > -\infty$ para todo valor de x. Una función convexa que no es propia es llamada* ***impropia****.*

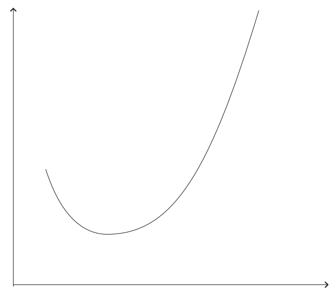

Figura 1.4: estríctamente convexa

Definición 1.6. *Sea S un conjunto no vacío en $\mathbb{R}^n$ y sea $f : S \to \mathbb{R}$. Entonces, se dice que f es **diferenciable** en $\bar{x} \in intS$ si existe un vector $\nabla f(\bar{x})$, llamado **vector gradiente**, y una función $\alpha : \mathbb{R}^n \to \mathbb{R}$ tal que*

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^t(x - \bar{x}) + \|x - \bar{x}\|\alpha(\bar{x}; x - \bar{x}), \quad \textit{para cada } x \in S,$$

donde $\lim\limits_{x \to \bar{x}} \alpha(\bar{x}; x - \bar{x}) = 0$.

§1.2. El concepto de regularización.

La idea de la regularización surge de los problemas mal planteados. 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 planteado, 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 es reemplazar L por el operador $L + \lambda M$, donde $\lambda \in \mathbb{R}$ y $M : \mathbb{X} \to \mathbb{X}$ son tales que el problema

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

está bien planteado, para algún $\lambda > 0$. En tal caso tiene solución única f_λ y es de esperar que a medida que $\lambda \to 0$, f_λ proporcione alguna aproximación al problema inicial dado.

Este concepto es aplicado a problemas de optimización. Por ejemplo, si tomamos $\mathbb{X} = \mathbb{R}^n$ y $L = \nabla f$ donde $f : \mathbb{R}^n \to \mathbb{R}$ es una función convexa, el caso inicial se convierte en:

$$\nabla f = 0$$

ó equivalentemente:

$$\min_{x \in \mathbb{R}^n} f(x). \tag{1.1}$$

Supongamos que f esta acotado inferiormente y tomemos $g : \mathbb{R}^n \to \mathbb{R}$ estríctamente convexa y coersiva. Asi el problema inicial dado pudiera no tener solución o más de una solución pero el problema regularizado:

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda g(x). \tag{1.2}$$

Tiene solución única para cada $\lambda > 0$, porque la función $f + \lambda g$ es coersiva (usando el hecho de que f es acotada inferiormente) lo cual reduce el problema a un conjunto compacto de modo que se garantiza la existencia de soluciones y también la convexidad estricta implicando la unicidad de la solución.

Este problema regularizado tiene soluciones únicas $x(\lambda)$ y en algunas hipótesis razonables (incluyendo la existencia de soluciones de el problema inicial) se puede probar que cuando $\lambda \to 0^+$, $x(\lambda)$ existe y resuelve (1.1).
El problema de aproximar ésta regularización es que aunque es estríctamente convexa y coersiva, para cualquier $\lambda > 0$ por pequeño que sea, esta función se comporta casi como la f o en otras palabras, si el sistema $\nabla f(x) = 0$ esta mal condicionado, entonces el sistema $(\nabla f + \lambda g)$ podría estar mal condicionado cuando $\lambda \to 0$ a pesar del hecho de que éste tenga una solución única $\forall \lambda > 0$.

§1.3. Algoritmo de punto proximal para optimización en $\mathbb{R}^n$.

Definición 1.7. *Una sucesión $\{y^k\}_{k \in \mathbb{N}}$ en $\mathbb{R}^n$ es llamada **Fejér convergente** a un conjunto $U \subset \mathbb{R}^n$ con respecto a la norma euclidiana si:*

$$\|y^{k+1} - u\| \leq \|y^k - u\|, \quad \forall k \geq 0, \quad \forall u \in U. \tag{1.3}$$

Definición 1.8. *Una función $f : \mathbb{R}^n \to \mathbb{R}$ se dice que es* ***coersiva o coersitiva*** *si*

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

Proposición 1.1. *Si $\{y^k\}_{k\in\mathbb{N}}$ es Fejér convergente a $U \neq \emptyset$ entonces, $\{y^k\}_{k\in\mathbb{N}}$ esta acotada. Si y es un punto en la frontera de la sucesión a lo largo de U entonces,*

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

Demostración:

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

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

Esto implica que:

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

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

■

Teorema 1.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ convexa y continuamente diferenciable. Supongamos que el conjunto de minimizadores de f en $\mathbb{R}^n$ es no vacío. Entonces la sucesión $\{x^k\}$ generada por:*

$$x^0 \in \mathbb{R}^n \tag{1.4}$$

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^n} \{f(x) + \lambda_k \|x - x^k\|^2\} \tag{1.5}$$

converge a un punto $x^* \in U$.
donde λ_k son números reales que satisfacen:

$$0 < \lambda_k \leqslant \bar{\lambda} \quad \forall k \in \mathbb{N}. \tag{1.6}$$

para algún $\bar{\lambda} > 0$ *y* $\|.\|$ *es la norma euclídea.*

Demostración:

La demostración se divide en 4 pasos. En el paso 1 se probará que $\{x^k\}$ esta bien definida. En segunda instancia veremos que $\{x^k\}$ es Fejér convergente a U. En el paso 3 se establecerá que el factor $\lim\limits_{k\to\infty}(x^{k+1} - x^k) = 0$, la cual se usará en el paso 4, donde se probará que el conjunto de puntos de la sucesión $\{x^k\}$ pertenece a U. De esta manera con los pasos desde el 2 hasta el 4, junto con la proposición (1.1), se obtiene la tesis del teorema.

Paso 1 La sucesión $\{x^k\}$ está bien definida.

Sea $f_k = f(x) + \lambda_k\|x - x^k\|^2$. Asi f alcanza un mínimo, pues esta es acotada inferiormente, además que $\lim\limits_{\|x\|\to\infty} f_k(x) = \infty$. Luego como f es contínua y la minimización en (1.5) reduce el problema a un conjunto compacto, se garantiza así la solución y por lo tanto f tiene un mínimo. Por otra parte, como f es convexa y además como $\lambda_k\|x - x^k\|^2$ es estríctamente convexa, entonces f_k es estríctamente convexa y de esta manera se garantiza que el mínimo es único y que $\{x^{k+1}\}$ son únicos, y por tanto, la sucesión esta bien definida.

Paso 2. Veamos que: $\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2,\ \forall k \geq 0$ *y todo* $\bar{x} \in U$.

$$\|x^k - \bar{x}\|^2 = \|x^k - x^{k+1} + x^{k+1} - \bar{x}\|^2$$

$$= \|x^k - x^{k+1}\|^2 + \|x^{k+1} - \bar{x}\|^2 + 2\langle x^k - x^{k+1}, x^{k+1} - \bar{x}\rangle. \tag{1.7}$$

Luego, como x^{k+1} resuelve (1.5) se tiene:

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

De donde

$$2(x^{k+1} - x^k) = -\frac{1}{\lambda_k}\nabla f(x^{k+1}). \tag{1.8}$$

Asi, de (1.7) y (1.8), se tiene que:

$$\begin{aligned}\|x^k-\bar{x}\|^2-\|x^{k+1}-x^k\|^2-\|x^{k+1}-\bar{x}\|^2 &= 2\langle x^k-x^{k+1},x^{k+1}-\bar{x}\rangle\\ &= \frac{1}{\lambda_k}\langle\nabla f(x^{k+1}),x^{k+1}-\bar{x}\rangle\\ &\geq \frac{1}{\lambda_k}[f(x^{k+1})-f(\bar{x})]\geq 0. \qquad (1.9)\end{aligned}$$

Ya que $\bar{x}$ es el minimizador de f. Por tanto, por la última desigualdad (1.9), obtenemos lo deseado.

Paso 3 Veamos ahora que:

$$\lim_{k\to\infty}(x^{k+1}-x^k)=0. \qquad (1.10)$$

Del caso enterior

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

Así, $\{\|x^k-\bar{x}\|\}$ es decreciente, y es acotada pues al ser decreciente

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

Por tanto tenemos que converge.
Para cada $k\geq 0$

$$\|x^{k+1}-x^k\|^2=\|x^k-\bar{x}\|^2-\|x^{k+1}-\bar{x}\|^2.$$

y así

$$\lim_{k\to+\infty}\|x^{k+1}-x^k\|^2=\lim_{k\to+\infty}\left(\|x^k-\bar{x}\|^2-\|x^{k+1}-\bar{x}\|^2\right)=0.$$

Luego, $\lim_{k\to+\infty}\|x^{k+1}-x\|^2=0$.

Paso 4. $\{x^k\}$ tiene puntos frontera y todos ellos pertenecen a U.

La existencia de los puntos fronteras están garantizados por el paso 2, siendo mas explicitos, se puede observar que $\{x^k\}$ es Fejér convergente a un conjunto U y por la primera parte de la proposición (1.1) tenemos que $\{x^k\}$ es acotada y asi queda garantizada la existencia de puntos frontera de la sucesión. Sea $\bar{x}$ un punto clausura de la sucesión $\{x^k\}$ y tomemos $\{x^{f_k}\}$ una subsucesión de $\{x^k\}$, tal que $\lim\limits_{k\to\infty} x^{f_k} = \bar{x}$. Por (1.8):

$$\nabla f(x^{f_k+1}) = 2\lambda_{j_k}(x^{f_k} - x^{f_k+1}). \tag{1.11}$$

Luego, por el caso 3 $\lim\limits_{k\to\infty} x^{f_k+1} = \lim\limits_{k\to\infty} x^{f_k} = \bar{x}$. Así, tomando límite en la igualdad (1.11) cuando $k \to \infty$, usando $\lambda_k \leq \bar{\lambda}$ y además que f es contínuamente diferenciable, tendríamos $\nabla f(\bar{x}) = 0$. Por la convexidad de f, $\bar{x} \in U$.

De esta forma por los casos 2,4 y la proposición (1.1) segunda parte, existe $x^* \in U$ tal que

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

. ■

Definición 1.9. *Diremos que ξ es un **subgradiente** de f en x si*

$$\langle \xi, y - x \rangle \leq f(y) - f(x), \quad \forall y \in \mathbb{R}^n. \tag{1.12}$$

Denotaremos por $\partial f(x)$ el conjunto de subgradientes de f en x, es decir,

$$\partial f(x) = \{\xi : \xi \text{ es un subgradiente de } f \text{ en } x\}.$$

.

§1.4. Funciones y distancias Bregman.

Definición 1.10. *Sea S un subconjunto abierto y convexo de $\mathbb{R}^n$ y $\bar{S}$ su clausura. Considere una función h convexa a valores reales definida en $\bar{S}$ y sea $D_h : \bar{S} \times S \to \mathbb{R}$ dada por:*

$$D_h(x, y) = h(x) - h(y) - \nabla h(y)^t(x - y). \tag{1.13}$$

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

1. *h es continuamente diferenciable en S.*

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

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

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* $\{x^k\}$ *esta acotada,* $\lim\limits_{k\to\infty} y^k = y^*$ *y* $\lim\limits_{k\to\infty} D_h(x^k, y^k) = 0$*, entonces* $\lim\limits_{k\to\infty} x^k = y^*$*.*

Donde D_h *es dada como en (1.13) y es llamada* ***distancia de Bregman*** *inducida por* h *y además a* S *se le conoce como la* ***zona*** *de* h*.*

Notas

1. $D_h(x,y) \geq 0$ para todo $x \in \bar{S}$ y $y \in S$. Para garantizar tal afirmación, es suficiente ver que h es estríctamente convexa y continuamente diferenciable en S, de esta manera satisface que:

$$h(x) \geq h(y) + \nabla h(y)^t(x-y), \ \ \forall x. \tag{1.14}$$

2. $D_h(x,y) = 0$ si y sólo si x=y.

Definición 1.11. *Vamos a introducir dos subclases que serán usadas más adelante.*

1. *Una función Bregman es llamada de* ***borde coercitivo****, 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 coercitiva****, si para cada* $y \in \mathbb{R}^n$ *existe un* $x \in S$ *tal que* $\nabla h(x) = y$*.*

veamos algunos ejemplos de funciones Bregman.

Ejemplo 1.1. $S = \mathbb{R}^n$, $h(x) = x^tMx$*, con* $M \in \mathbb{R}^{n\times n}$ *simétrica y definida positiva. En este caso* $D_h(x,y) = (x-y)^tM(x-y) = \|x-y\|_M^2$*.*

En efecto:

h es una función cuadrática y asi sabemos que es continuamente diferenciable en $\mathbb{R}^n$ y además, por ser definida positiva la matriz M, nos dice que la función es estríctamente convexa en $\mathbb{R}^n$. Calculemos la distancia D_h.
$D_h(x,y) = h(x) - h(y) - \nabla h(y)^t(x-y)$, luego 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 1.2. *$S = \mathbb{R}^n_{++}$, $h(x) = \sum_{j=1}^{n} \log(x_j)$ extendido con continuidad en la frontera usando la convención de que $0\log(0) = 0$.En este caso , $D_h(x,y) = \sum_{j=1}^{n}\left(x_j \log\left(\frac{x_j}{y_j}\right) + y_j - x_j\right)$. Esta función es la llamada **Kullback-Leiber**.*

Ejemplo 1.3. *Sea $S = \mathbb{R}^n_{++}$, $h(x) = \sum_{i=1}^{n}(x_i^\alpha - x_i^\beta)$, con $\alpha \geq 1$, $0 < \beta < 1$. Para $\alpha = 2$ y $\beta = \frac{1}{2}$ se tiene:*

$$D_h(x,y) = \|x-y\|^2 + \sum_{i=1}^{n} \frac{1}{2\sqrt{y_i}}(\sqrt{x_i} - \sqrt{y_i})^2$$

y para $\alpha = 1$, $\beta = \frac{1}{2}$ se tiene:

$$d_h(x,y) = \sum_{i=1}^{n} \frac{1}{2\sqrt{y_i}}(\sqrt{x_i} - \sqrt{y_i})^2.$$

Proposición 1.2. *Si h es una función Bregman con zona S, entonces:*

(i) $D_h(x,y) - D_h(x,z) - D_h(z,y) = \langle \nabla h(y) - \nabla h(z), z - x\rangle$.

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

(iii) $D_h(.,y)$ *es estríctamente 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.

para (ii) es trivial, basta aplicar gradiente respecto a x a (i).

(iii) Sea y fijo pero arbitrario en S y consideremos la función de una variable $f(x) = D(x,y) = h(x) - h(y) - \nabla h(y)^t(x-y)$ 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 estríctamente convexa y asi $D_h(.,y)$ es estríctamente convexa, para todo $y \in S$.

■

§1.5. El método de punto proximal para distancias Bregman.

El problema en interés es:

$$\begin{aligned} &\min f(x) \\ &s.a.\ \ x \in \bar{S}. \end{aligned} \tag{1.15}$$

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

$$x^0 \in S. \tag{1.16}$$

$$x^{k+1} = arg \min_{x \in \bar{S}} \{f(x) + \lambda_k D_h(x, x^k)\} \tag{1.17}$$

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

$$0 < \lambda_k \leq \bar{\lambda}, \tag{1.18}$$

para algún $\bar{\lambda} > 0$. Estudiemos ahora la convergencia de dicha sucesión. Asumiremos que (1.15) tiene solución y además que f es acotada inferiormente.

Teorema 1.2. *Si el problema (1.15) tiene solución y h es de borde coercivo con respecto a S, entonces la sucesión $\{x^k\}$ generada por (1.16) y (1.17) converge a una solución x^* del problema (1.15).*

Demostración:

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

Sea β una cota inferior para f en $\bar{S}$ y sea $f_k(x) = f(x) + \lambda_k D_k(x, x^k)$. Entonces $f_k(x) \geq \beta + \lambda_k D_k(x, x^k)$ y así por definición 1.10(3) sigue que el conjunto de niveles de f_k es acotado, luego la minimización en 1.17 reduce el problema a un conjunto compacto y asi el mínimo está logrado. Por otro lado, f_k es estríctamente convexo, pues $D_k(\cdot, x^k)$ es estríctamente convexa por proposición 1.2(iii), asi el mínimo es único y $\{x^k\}$ estará bien definida.

Veamos ahora que la sucesión $\{x^{k+1}\} \subset S$. Es fácil ver para (1.17) que $\{x^{k+1}\}$ es el único $x \in \bar{S}$ tal que:

$$\lambda_k \nabla h(x^k) \in \partial(f + \lambda_k h)(x). \tag{1.19}$$

Se probará bajo la definición 1.11(1), $\partial(f + \lambda_k h)(x) = \emptyset$ lo cual implica, en vista de (1.19), que $\{x^{k+1}\} \subset S$. Tomemos $x \in \partial S$ y supongamos que existe $\xi \in \partial(f+\lambda_k h)(x)$, además $z \in S$ y definamos:

$$y^\ell = (1 - \varepsilon_\ell)x + \varepsilon_\ell z, \ \textit{donde} \ \lim_{\ell\to\infty} \varepsilon_\ell = 0 \ y \ \varepsilon_\ell \neq 1. \tag{1.20}$$

Trabajando (1.20), podemos obtener:

$$y^\ell = (1 - \varepsilon_\ell)x + \varepsilon_\ell z \Longleftrightarrow y^\ell - x = \varepsilon_\ell(z - x). \tag{1.21}$$

Luego por la definición (1.20) tenemos que $y^\ell \in S$, por la convexidad de S y $\lim\limits_{\ell\to\infty} y^\ell = x$. Además:

$$\begin{aligned}
\varepsilon_\ell \xi^t(z - x) &= \xi^t(y^\ell - x) && (Por 1.21)\\
&\leq f(y^\ell) - f(x) + \lambda_k(h(y^\ell) - h(x)) && (def.1.12)\\
&\leq f(y^\ell) - f(x) + \lambda_k \nabla h(y^\ell)^t(y^\ell - x) && (diferenciabilidad\ h)\\
&\leq \varepsilon_\ell(f(z) - f(x)) + \lambda_k \frac{\varepsilon_\ell}{1-\varepsilon_\ell} \nabla h(y^\ell)^t(z - y^\ell). && (1.22)
\end{aligned}$$

Luego de esta última desigualdad, agrupando los términos:

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

Veamos (1.22). Como f es convexa en $\bar{S}$ tenemos que:

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

y además:

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

Luego, como $\lim_{\ell\to\infty} y^\ell = x \in \partial S$, por definición 1.11(1) implica que la parte derecha de (1.22) tiende a $-\infty$, cuando $\ell \to \infty$, por definición; pero la parte izquierda tiene límite finito y por tanto es una contradicción . Así, $\partial(f+\lambda_k h) = \emptyset, \forall x \in \partial S$ y además $\{x^{k+1}\} \in S$.

Paso 2. Veamos que $D_h(\bar{x}, x^{k+1}) \leq D_h(\bar{x}, x^k) - D_h(x^{k+1}, x^k)$ para todo k y cada solución $\bar{x}$ de (1.15).
Usando la proposición 1.2(i), con $x = \bar{x}, y = x^k, z = x^{k+1}$ se tiene que:

$$D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^{k+1} - \bar{x}\rangle. \quad (1.23)$$

Por (1.17)

$$0 \in \partial[f + \lambda D_h(\cdot, x^k)](x^{k+1}). \quad (1.24)$$

Ahora, por (1.23) y proposición (1.2)(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 (1.23) y la definición de subgradiente tenemos:

$$\begin{aligned}
D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) &= \langle \frac{1}{\lambda_k}(\lambda_k \nabla h(x^k) - \nabla h(x^{k+1})), x^{k+1} - \bar{x}\rangle \\
&= \frac{1}{\lambda_k}\langle y^k), x^{k+1} - \bar{x}\rangle \\
&\geq \frac{1}{\lambda_k}(f(x^{k+1}) - f(\bar{x})) \\
&\geq 0. \qquad\qquad (\bar{x}\, es\, minimizador\, de\, f\, en\, \overline{S})
\end{aligned} \tag{1.25}$$

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

Así,

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

Paso 3. La sucesión $\{x^k)\}$ esta acotada y $\lim_{k\to\infty} x^{f_k} = \hat{x}$ implica que $\lim_{k\to\infty} x^{f_k+1} = \hat{x}$. Por el paso 2 $\{D_k(\bar{x}, x^k)\}$, es decreciente y la no negatividad de $D_k(x, y)$ nos dice que la sucesión $\{D_h(\bar{x}, x^k)\}$ es convergente. Por la conclusión del paso 2, aplicando límite, tenemos:

$$\lim_{k\to\infty} D_h(x^{k+1}, x^k) = 0. \tag{1.26}$$

Por otro lado, como $\{D_h(\bar{x}, x^k)\}$ es decreciente, obtenemos asi que la sucesión es acotada pues $D_h(\bar{x}, x^k) \leq D_h(\bar{x}, x^0)$.

$\therefore$ Por definición 1.10(3) $\{x^k\}$ esta acotada. Ahora si consideramos, $\lim_{k\to\infty} x^{f_k} = \hat{x}$, donde $\{x^{f_k}\}$ es una subsucesión de x^k, por definición 1.10(5) se puede concluir que:

$$\lim_{k\to\infty} x^{f_k+1} = \hat{x},$$

por (1.23) y el hecho de que sucesión x^{f_k+1} es también una subsucesión de x^k(acotada).

Paso 4. Todos los puntos frontera de $\{x^k\}$ son solución de (1.15).

Tomemos un punto $\bar{x}$ solución de (1.15). Sea $\hat{x}$ un punto frontera de $\{x^k\}$ y $\{x^{f_k+1}\}$ una subsucesión de $\{x^k\}$, tal que $\lim_{k\to\infty} x^{f_k} = \hat{x}$. La existencia de $\hat{x}$ sigue del *paso* 3.

$$
\begin{aligned}
0 &\leq \frac{1}{\bar{\lambda}}(f(x^{f_k+1}) - f(\bar{x})) \\
&\leq \frac{1}{\lambda_k}(f(x^{f_k+1}) - f(\bar{x})) \\
&\leq D_h(\bar{x}, x^{f_k}) - D_h(\bar{x}, x^{f_k+1}) - D_h(x^{f_k+1}, x^{f_k}) \longrightarrow_{k\to\infty} 0 \qquad (Por 1.25) \qquad (1.27)
\end{aligned}
$$

por la convergencia de $D_h(\bar{x}, x^k)$ y definición 1.10(5). Por (1.27) tomando límite cuando $k \to \infty$, $f(\hat{x}) = f(\bar{x})$. Luego, como $\overline{S}$ es cerrado y $\{x^k\} \subset \overline{S}$ se tiene que $\hat{x}$ resuelve (1.15).

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 $\hat{x}$ un punto clausura de $\{x^k\}$ y consideramos una subsucesión $\{x^{f_k}\}$ de $\{x^k\}$ tal que $\lim\limits_{k\to\infty} x^{f_k} = \hat{x}$. Por la definición 1.10(4) $\lim\limits_{k\to\infty} D_h(x^{f_k}, \hat{x}) = 0$. Pero por el paso 4, $\hat{x}$ resuelve el problema (1.15)y por el paso 2 $\{D_h(\hat{x}, x^k)\}$ es una sucesión no negativa y decreciente con una subsucesión convergiendo a cero. De aca sigue que la sucesión converge a cero y por definición 1.10(4) se tiene $\lim\limits_{k\to\infty} x^k = \hat{x}$.

■

Capítulo 2

$\varphi-$divergencia.

La mayoría de los conceptos y resultados expuestos acá serán usados en el análisis de convergencia del algoritmo, por tal razón hemos considerado que este capítulo es de suma importancia para este trabajo, por ejemplo, presentaremos la definiciones de funciones cuasiconvexas y $\varphi-$divergencia, además de algunas de sus propiedades básicas, las cuales son utilizadas en el contexto de la optimización. Esta clase de medida fue presentado por Csiszár [3] en 1967 como una medida de información generalizada. En el contexto de los métodos de punto proximal, la $\varphi-$divergencia fue estudiada inicialmente por Teboulle [11] en 1992, donde varias de sus propiedades son presentadas.

§2.1. Funciones cuasiconvexas.

Definición 2.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}$. Entonces, f es llamada **cuasiconvexa**, si para todo $x, y \in domf$ y $\lambda \in (0,1)$ se cumple que:*

$$f(\lambda x + (1-\lambda)y) \leq máx\{f(x), f(y)\}.$$

Definición 2.2. *sea $f : \mathbb{R}^n \to \mathbb{R}$. Entonces, f es llamada **estríctamente cuasiconvexa**, si para todo $x, y \in domf$ con $f(x) \neq f(y)$ se cumple que:*

$$f(\lambda x + (1-\lambda)y) < máx\{f(x), f(y)\} \; \forall \lambda \in (0,1).$$

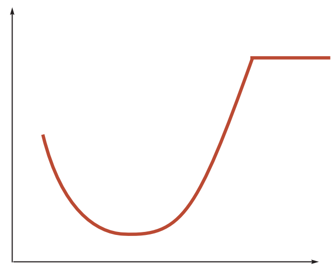

Figura 2.1: función cuasiconvexa.

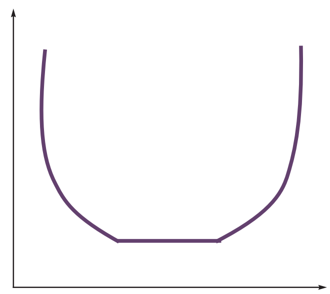

Figura 2.2: función estríctamente cuasiconvexa.

Definición 2.3. *Sea $f : \mathbb{R}^n \to \mathbb{R}$. Entonces, f es llamada* ***fuertemente cuasiconvexa*** *si para todo $x, y \in$ dom$f$ con $x \neq y$, entonces:*

$$f(\lambda x + (1-\lambda)y) < \text{máx}\{f(x), f(y)\} \ \forall \lambda \in (0,1).$$

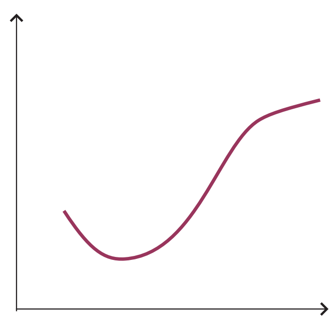

Figura 2.3: fuertemente cuasiconvexa pero no convexa.

Teorema 2.1. *Una función $f : \mathbb{R}^n \to \mathbb{R}$ es cuasiconvexa si y sólo si el conjunto de nivel $S_\alpha := \{x \in domf | f(x) \leq \alpha\}$ es convexo $\forall \alpha \in \mathbb{R}$.*

Demostración:

Supongamos que f es cuasiconvexa. Sea $\alpha \in \mathbb{R}$ cualquiera y sean $x_1, x_2 \in S_\alpha$ asi que $f(x_1) \leq \alpha$ y $f(x_2) \leq \alpha$. Sea $\lambda \in (0,1)$ y sea $x = \lambda x_1 + (1-\lambda)x_2$. Luego por la cuasiconvexidad de f, $f(x) \leq \text{máx}\{f(x_1), f(x_2)\} \leq \alpha$. Así, $x \in S_\alpha$ y S_α es convexo. Recíprocamente, supongamos que S_α es convexo para cada $\alpha \in \mathbb{R}$. Sean $x_1, x_2 \in dom(f)$ y $\lambda \in (0,1)$. Haciendo $\alpha = \text{máx}\{f(x_1), f(x_2)\}$ tenemos que $x_1, x_2 \in S_\alpha$ y por la convexidad de S_α se tiene que $\lambda x_1 + (1-\lambda)x_2 \in S_\alpha$, es decir

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \ \alpha = \text{máx}\{f(x_1), f(x_2)\}.$$

Por lo tanto f es cuasiconvexa.

■

Teorema 2.2. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función diferenciable. Entonces, f es cuasiconvexa si y sólo si $\langle \nabla f(y), y - x \rangle \geq 0$ siempre que $f(x) \leq f(y)$, $\forall x, y \in \mathbb{R}^n$.*

Demostración:

Sea f una función cuasiconvexa y sea $x, y \in \mathbb{R}^n$ tal que $f(x) \leq f(y)$. Por la diferenciabilidad de f en y, y para $\lambda \in (0,1)$ se tiene que:

$$f(\bar{x}) = f(y) + \nabla f(y)^t(\bar{x} - y) + \|\bar{x} - y\|\alpha(y; \bar{x} - y) \quad \forall \bar{x} \in \mathbb{R}^n.$$

donde $\alpha(y; \bar{x} - y) \to 0$ cuando $\bar{x} \to y$

Tomando $\bar{x} = y - \lambda(y - x)$

$$\Rightarrow f(y - \lambda(y-x)) = f(y) + \nabla f(y)^t((y - \lambda(y-x)) - y) + \|(y - \lambda(y-x)) - y\| + \alpha(y; (y - \lambda(y-x)$$
$$\Rightarrow f(y - \lambda y + \lambda x) = f(y) + \nabla f(y)^t(y - \lambda y + \lambda x - y) + \|y - \lambda y + \lambda x - y\| + \alpha(y; y - \lambda y + \lambda x - y)$$
$$\Rightarrow f(\lambda x + (1-\lambda)y) = f(y) + \nabla f(y)^t\lambda(x - y) + |\lambda|\|x - y\|\alpha(y; \lambda(x-y)).$$

esto implica

$$f(\lambda x + (1-\lambda)y) - f(y) = \lambda \nabla f(y)^t(x - y) + \lambda\|x - y\|\alpha(y; \lambda(x-y)). \qquad (2.1)$$

Por la cuasiconvexidad de f, tenemos que:
Para todo $x, y \in \mathbb{R}$ y $\lambda \in (0,1)$.

$$\begin{aligned} f(\lambda x+(1-\lambda)y) &\leq m\acute{a}x\{f(x),f(y)\} \\ &= f(y). \end{aligned}$$

De acá,

$$f(\lambda x+(1-\lambda)y)-f(y)\leq 0 \tag{2.2}$$

De (3.1) y (3.7), obtenemos:

$$\lambda\nabla f(y)^t(x-y)+\lambda\|x-y\|\alpha(y;\lambda(x-y))\leq 0.$$

Dividiendo entre λ, con $\lambda\in(0,1)$

$$\nabla f(y)^t(x-y)+\|x-y\|\alpha(y;\lambda(x-y))\leq 0. \tag{2.3}$$

Aplicando límite cuando $\lambda\to 0$, se tiene:

$$\begin{aligned} &\nabla f(y)^t(x-y)\leq 0,\ \text{ puesto que } \lim_{\lambda\to 0}\alpha(y;\lambda(x-y))=0 \\ \Rightarrow\ &\nabla f(y^t)(y-x)\geq 0. \end{aligned}$$

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

Supongamos que para todo $x,y\in\mathbb{R}^n$ se tiene que $f(x)\leq f(y)$ implica que $\langle\nabla f(x),y-x\rangle\geq 0$. Queremos probar que f es cuasiconvexa. Sean $x,y\in\mathbb{R}$ cualesquiera.
Sea $L=\{\bar{x}:\bar{x}=\lambda x+(1-\lambda)y$, para algún $\lambda\in(0,1)$ y $f(\bar{x})>f(y)\}$. Probemos que L es vacío, usando reducción al absurdo.
Tomemos $\tilde{x}\in L$.
Por tanto, $\tilde{x}=\lambda x+(1-\lambda)y$ para algún $\lambda\in(0,1)$ y $f(\tilde{x})>f(y)$. Como f es diferenciable, entonces f es continua. Sea $\gamma=\|\tilde{x}-y\|$. Entonces, existe $\epsilon\in(0,\gamma)$ tal que $f(x)>f(y)$, para todo $x\in B_\epsilon(\tilde{x})$.

Afirmacion 1. $x_\epsilon=(1-\frac{\epsilon}{\gamma})\tilde{x}+\frac{\epsilon}{\gamma}y\in\partial B_\epsilon(\tilde{x})$.

En efecto,

$$\begin{aligned}
x_\epsilon = \left(1-\frac{\epsilon}{\gamma}\right)\tilde{x} + \frac{\epsilon}{\gamma}y &\Rightarrow x_\epsilon = \tilde{x} - \frac{\epsilon}{\gamma}\tilde{x} + \frac{\epsilon}{\gamma}y \\
&\Rightarrow x_\epsilon = \tilde{x} + \frac{\epsilon}{\gamma}(y-\tilde{x}) \\
&\Rightarrow x_\epsilon - \tilde{x} = \frac{\epsilon}{\gamma}(y-\tilde{x}) \\
&\Rightarrow \|x_\epsilon - \tilde{x}\| = \left\|\frac{\epsilon}{\gamma}(y-\tilde{x})\right\| \\
&\Rightarrow \|x_\epsilon - \tilde{x}\| = \frac{\epsilon}{\gamma}\|y-\tilde{x}\| \\
&\Rightarrow \|x_\epsilon - \tilde{x}\| = \frac{\epsilon}{\gamma}\gamma \\
&\Rightarrow \|x_\epsilon - \tilde{x}\| = \epsilon.
\end{aligned}$$

Supongamos que $f(x_\epsilon) \geq f(\tilde{x})$, para todo $\epsilon \in (0,\gamma)$.
Luego,

$$\lim_{\epsilon\to\gamma} f(x_\epsilon) \geq f(\tilde{x}) \Rightarrow f(y) \geq f(\tilde{x}), \quad \text{ya que } \lim_{\epsilon\to\gamma} f(x_\epsilon) = f(y).$$

Lo cual contradice que $f(\tilde{x}) > f(y)$. Así, existe $\epsilon \in (0,\gamma)$ tal que $f(x) > f(y)$, para todo $x \in B_\epsilon(\tilde{x})$ y $f(x_\epsilon) < f(\tilde{x})$.

Por otro lado, haciendo $\delta = 1 - \frac{\epsilon}{\gamma}$

$$f(x_\epsilon) = f(\delta\tilde{x} + (1-\delta)y) < f(\tilde{x}). \tag{2.4}$$

Ahora , probemos que $\forall\mu \in (\delta,1], x_\mu = \mu\tilde{x} + (1-\mu)y \in B_\epsilon(\tilde{x})$.

$$\begin{aligned}
x_\mu - \tilde{x} = \mu\tilde{x} + (1-\mu)y - \tilde{x} &\Rightarrow x_\mu - \tilde{x} = (\mu-1)\tilde{x} + (1-\mu)y \\
&\Rightarrow x_\mu - \tilde{x} = (1-\mu)(y-\tilde{x}) \\
&\Rightarrow \|x_\mu - \tilde{x}\| = \|(1-\mu)(y-\tilde{x})\| \\
&\Rightarrow \|x_\mu - \tilde{x}\| = (1-\mu)\|y-\tilde{x}\| \\
&\Rightarrow \|x_\mu - \tilde{x}\| = (1-\mu)\gamma < (1-\delta)\gamma, \quad (\delta < \mu \leq 1) \\
&\Rightarrow \|x_\mu - \tilde{x}\| < \frac{\epsilon}{\gamma}\gamma \\
&\Rightarrow \|x_\mu - \tilde{x}\| < \epsilon.
\end{aligned}$$

Por lo tanto, $\forall\mu \in (\delta,1]$, $x_\mu \in B_\epsilon(\tilde{x})$. Así, que

$$f(x_\mu) > f(y), \text{ para todo } \mu \in (\delta,1]. \tag{2.5}$$

Por la inecuación (2.4) y el teorema de valor medio, se tiene:

$$0 < f(\tilde{x}) - f(\delta\tilde{x} + (1-\delta)y) = (1-\delta)\nabla f(\hat{x})^t(\tilde{x} - y), \tag{2.6}$$

donde $\hat{x} = \hat{\mu}\tilde{x} + (1-\hat{\mu})$ para algún $\hat{\mu} \in (\delta, 1]$.
De (2.5), tenemos que $f(\hat{x}) > f(y)$.
Dividiendo (2.6) por $1-\delta > 0$, se deduce que $\nabla f(\hat{x})^t(\tilde{x}-y) > 0$, que a su vez implica que:

$$\begin{aligned}
\nabla f(\hat{x})^t(\hat{\mu}\tilde{x} + (1-\hat{\mu})y - y) > 0 &\Rightarrow \nabla f(\hat{x})^t(\hat{\mu}\tilde{x} + y - \hat{\mu}y - y) > 0 \\
&\Rightarrow \nabla f(\hat{x})^t(\hat{\mu}(\tilde{x} - y)) > 0 \\
&\Rightarrow \hat{\mu}\nabla f(\hat{x})^t(\tilde{x} - y) > 0 \\
&\Rightarrow \nabla f(\hat{x})^t(\tilde{x} - y) > 0, \qquad (\hat{\mu} \in (\delta, 1]) \\
&\Rightarrow \nabla f(\hat{x})^t(\lambda x + (1-\lambda)y - y) > 0, \qquad (\lambda \in (0, 1)) \\
&\Rightarrow \nabla f(\hat{x})^t(\lambda x + y - \lambda y - y) > 0 \\
&\Rightarrow \nabla f(\hat{x})^t(\lambda(x - y)) > 0 \\
&\Rightarrow \lambda\nabla f(\hat{x})^t(x - y) > 0 \\
&\Rightarrow \nabla f(\hat{x})^t(x - y) > 0. \qquad (2.7)
\end{aligned}$$

Por otra parte, $f(\hat{x}) > f(y) \geq f(x)$.
Veamos que $\hat{x}$ es una combinación convexa de x y y. En efecto,

$$\begin{aligned}
\hat{x} = \hat{\mu}\tilde{x} + (1-\hat{\mu})y &\Rightarrow \hat{x} = \hat{\mu}(\lambda x + (1-\lambda)y) + (1-\hat{\mu})y \\
&\Rightarrow \hat{x} = \hat{\mu}(\lambda(x - y) + y) + y - \hat{\mu}y \\
&\Rightarrow \hat{x} = \hat{\mu}\lambda(x - y) + \hat{\mu}y + y - \hat{\mu}y \\
&\Rightarrow \hat{x} = \hat{\mu}\lambda x - \hat{\mu}\lambda y + y \\
&\Rightarrow \hat{x} = \hat{\mu}\lambda x + (1 - \hat{\mu}\lambda)y \\
&\Rightarrow \hat{x} = \hat{\lambda}x + (1-\hat{\lambda})y, \qquad (\text{haciendo } \hat{\lambda} = \hat{\mu}\lambda \in (0, 1)).
\end{aligned}$$

Por hipótesis y dado que $f(\hat{x}) > f(x)$, se tiene que $\nabla f(\hat{x})^t(x - \hat{x}) \leq 0$ y de esta manera se obtiene:

$$\begin{aligned}
0 \geq \nabla f(\hat{x})^t(x-\hat{x}) &\Rightarrow 0 \geq \nabla f(\hat{x})^t(x-(\hat{\mu}\hat{x}+(1-\hat{\mu})y)) \\
&\Rightarrow 0 \geq \nabla f(\hat{x})^t(x-(\hat{\mu}(\lambda x+(1-\lambda)y)+y-\hat{\mu}y)) \\
&\Rightarrow 0 \geq \nabla f(\hat{x})^t(x-\hat{\mu}\lambda x-\hat{\mu}y+\hat{\mu}\lambda y-y+\hat{\mu}y) \\
&\Rightarrow 0 \geq \nabla f(\hat{x})^t((1-\hat{\mu}\lambda)x-(1-\hat{\mu}\lambda)y) \\
&\Rightarrow 0 \geq \nabla f(\hat{x})^t((1-\hat{\mu}\lambda)(x-y)) \\
&\Rightarrow 0 \geq (1-\hat{\mu}\lambda)\nabla f(\hat{x})^t(x-y) \\
&\Rightarrow 0 \geq (1-\hat{\lambda})\nabla f(\hat{x})^t(x-y), \quad \hat{\lambda}=\hat{\mu}\lambda \in (0,1) \\
&\Rightarrow 0 \geq \nabla f(\hat{x})^t(x-y). \qquad (2.8)
\end{aligned}$$

Pero la inecuación anterior contradice a (2.7).
Así L es vacío.
Por lo tanto, $f(\lambda x+(1-\lambda)y) \leq f(y) = máx\{f(x), f(y)\}$, $\forall x, y \in \mathbb{R}^n$.
Por consiguiente, f es cuasiconvexa.
■

§2.2. $\varphi-$divergencia

Sea $\varphi : \mathbb{R}_{++} \to \mathbb{R}$ una función propia, convexa y cerrada que sastifacen las siguientes condiciones:

(Φ1) φ es dos veces continuamente diferenciable en $int(dom\varphi) = (0, +\infty)$.

(Φ2) φ es estríctamente convexa en este dominio.

(Φ3) $\varphi(1){=}\varphi'(1) = 0$ y $\varphi''(1) > 0$.

(Φ4) $\lim\limits_{t\to 0^+} \varphi'(t) = -\infty$.

Esta clase de funciones son de mucha utilidad en la definición de las $\varphi-$divergencias.

Definición 2.4. *En el resto de este trabajo se denotará por* Φ *la clase de funciones que satisfaga* (Φ1) $-$ (Φ4). *Definamos ahora dos subclases de* Φ.

$$\Phi_1 = \{\varphi \in \Phi : \varphi'(t) \leq \varphi''(1)\ln(t), \quad \forall t > 0\} \qquad (2.9)$$

y

$$\Phi_2 = \left\{\varphi \in \Phi : \varphi''(1)\left(1-\frac{1}{t}\right) \leq \varphi'(t) \leq \varphi''(1)\ln(t), \quad \forall t > 0\right\}. \tag{2.10}$$

Observación 2.1. *Como sabemos que $ln(t) \leq t-1$ para todo $t > 0$ y como por definición (2.4) $\varphi''(1) > 0$, así se tiene que $\Phi_2 \subset \Phi_1 \subset \Phi$.*

Definición 2.5. *Si $\varphi \in \Phi$, entonces $d_\varphi : \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \Rightarrow \mathbb{R}$ definida por:*

$$d_\varphi(x,y) = \sum_{i=1}^{n} y_i \varphi\left(\frac{x_i}{y_i}\right), \tag{2.11}$$

es llamado $\varphi-divergencia$.

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 $\varphi_i, i = \{1,2\}$ están en Φ_2, pues $\varphi_1'(t) = ln(t)$ y asi, como $\varphi_1''(t) = \dfrac{1}{t}$ entonces $\varphi_1''(1) = 1$, por tanto se tiene que:

$(1-\dfrac{1}{t}) \leq \varphi'' \leq \varphi''(1)\ln(t) = ln(t), \quad \forall t > 0$, asi se tiene que $\varphi_1 \in \Phi_2$. De manera análoga podemos ver que $\varphi_2'(t) = -\dfrac{1}{t} + 1 = 1 - \dfrac{1}{t}$ con lo cual sabiendo que $\varphi_2''(1) = 1$ se tiene

$1-\dfrac{1}{t} = 1-\dfrac{1}{t} \leq ln(t)$ y de esta manera $\varphi_2 \in \Phi_2$. Más aún por la observación (2.1) se tiene que $\varphi_1, \varphi_2 \in \Phi_1$.

La función φ_1 juega un papel importante en el análisis de convergencia que será estudiado en sección 3.2. Esta función φ_1 genera una cuasidistancia llamada también $kullback-Leibler$ o distancia entrópica y es dada por:

$$H(x,y) = d_\varphi(x,y) = \sum_{j=1}^{n} x_j ln\left(\frac{x_j}{y_j}\right) + y_j - x_j. \tag{2.12}$$

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

Observación 2.2. *Nuestra atención estará centrada en:*

$$\varphi(t) := \varphi_2(t) = t - \log t - 1. \tag{2.13}$$

En este caso,

$$d_\varphi(x,y) := d(x,y) = \sum_{i=1}^{n}\left(y_i ln\left(\frac{y_i}{x_i}\right) + x_i - y_i\right), \tag{2.14}$$

Donde el dominio puede ser extendido continuamente a $\mathbb{R}^n_{++} \times \mathbb{R}^n_+$, usando como convención que $0\ln(0) = 0$. En otras palabras, d admite a cero en su segundo argumento.

Proposición 2.1. *Sea $\varphi \in \Phi$, entonces*

(a) $\varphi(t) \geq 0$ *y* $\varphi(t) = 0$ *si y sólo si* $t = 1$*;*

(b) $\varphi(t)$ *es decreciente en* $(0,1)$ *y creciente en* $(1,+\infty)$*;*

(c) $\lim_{t\to+\infty} \varphi(t) = +\infty$*;*

(d) *Si* $\varphi(t) := \varphi_2(t) = t - \log t - 1$*, entonces* $\lim_{t\to 0^+} \varphi(t) = +\infty$*.*

Demostración:
Sea $\varphi \in \Phi$ cualquiera. Por ($\Phi3$) se tiene que $\varphi(1) = 0$ es un mínimo local de φ y junto con ($\Phi2$) tenemos que $\varphi(1)$ es un mínimo global de φ, así que (a) y (b) se cumplen pues $dom(\varphi) = (0,+\infty]$. Por (b) y por ($\Phi2$) se obtiene (c), la parte (d) es directa. ■

Lema 2.3. *Sea d definida como en (2.14). Entonces se tienen los siguientes resultados:*

(a) *El conjunto de nivel de* $d(\cdot,y)$ *es acotado para todo* $y \in \mathbb{R}^n_+$*;*

(b) *Si* $\{y^k\} \subset \mathbb{R}^n_{++}$ *converge a* $y \in \mathbb{R}^n_+$*, entonces* $\lim_{k\to+\infty} d(y^k,y) = 0$*;*

(c) *Si* $\{z^k\} \subset \mathbb{R}^n_+$*,*$\{y^k\} \subset \mathbb{R}^n_{++}$ *son sucesiones tales que* $\{z^k\}$ *esta acotada,* $\lim_{k\to+\infty} y^k = y$*, y además* $\lim_{k\to+\infty} d(y^k,z^k) = 0$*, entonces* $\lim_{k\to+\infty} z^k = y$*.*

Demostración:

Parte **(a)**. Tomemos:

$$\Gamma_1(y,\delta) = \{x \in \mathbb{R}^n_+ : d(x,y) \leq \delta\} \tag{2.15}$$
$$= \left\{ x \in \mathbb{R}^n_+ : \sum_{i=1}^{n} \left(y_i ln\left(\frac{y_i}{x_i}\right) + x_i - y_i \right) \leq \delta \right\}. \tag{2.16}$$
$$\tag{2.17}$$

Supongamos que Γ_1 no es acotado.

Luego, existe $\{x_n\} \subset \mathbb{R}^n_+$ tendiendo a $+\infty$, tal que:

$$y \ln\left(\frac{y}{x_n}\right) + x_n - y \leq \delta, \quad \forall n.$$

Lo que implica que,

$$x_n\left(\frac{y}{x_n}\ln\left(\frac{y}{x_n}\right) + 1\right) - y \leq \delta \tag{2.18}$$

pero $x_n\left(\frac{y}{x_n}\ln\left(\frac{y}{x_n}\right)+1\right) - y \to +\infty$ cuando $n \to +\infty$.

Lo que contradice (2.16)

así, Γ_1 es acotado para todo $y \in \mathbb{R}^n_+$.

Parte **(b)**. Usando la hipótesis y la definición de d se tiene:

$$\lim_{k\to+\infty} d(y,y^k) = \lim_{k\to+\infty} \sum_{i=1}^{n} \left(y_i ln\left(\frac{y_i}{y_i^k}\right) + y_i^k - y_i \right)$$
$$= \sum_{i=1}^{n} \lim_{k\to+\infty} \left(y_i ln\left(\frac{y_i}{y_i^k}\right) + y_i^k - y_i \right).$$

Si $y_i = 0$, entonces

$$y_i \ln\left(\frac{y_i}{y_i^k}\right) + y_i^k - y_i = 0\ln(0) + y_i^k = y_i^k$$

y así

$$\lim_{k\to+\infty}\left[y_i \ln\left(\frac{y_i}{y_i^k}\right) + y_i^k - y_i \right] = y_i = 0.$$

Si $y_i \neq 0$, entonces

$$y_i ln\left(\frac{y_i}{y_i}\right) + y_i - y_i = y_i ln(1) + y_i - y_i = y_i 0 + 0 = 0.$$

Por lo tanto, $\lim\limits_{k\to+\infty} d(y,y^k)=0$.

Parte **(c)**. Sea $\{z^k\}\subset\mathbb{R}^n_+$ acotado. Podemos escoger una subsucesión de $\{z^k\}$ convergente, digamos $\{z^{kj}\}$ con $\{z^{kj}\}\to\hat{z}$.
Sea $i\in\{1,...,n\}$ cualquiera.
Si $y_i^k\to y_i=0$, debemos probar que $z_i^{kj}\to 0$. Supongamos que $z_i^{kj}\nrightarrow 0$, es decir, $\hat{z}_i>0$.
Así que

$$\lim_{j\to+\infty}\left(z_i^{kj}\ln\left(\frac{z_i^{kj}}{y_i^{kj}}\right)+y_i^{kj}-z_i^{kj}\right)=-\infty$$

lo cual contradice que $\lim\limits_{k\to+\infty} d(y^k,z^k)=\lim\limits_{k\to+\infty}\sum\limits_{k\to+\infty}\left(z_i^{kj}\ln\left(\frac{z_i^{kj}}{y_i^{kj}}\right)+y_i^{kj}-z_i^{kj}\right)=0.$
Por lo tanto $z_i^{kj}\to 0$ o bien $\hat{z}_i=0$ si $y_i=0$.
Si $y_i^k\to y_i>0$, entonces

$$0=\lim_{j\to+\infty}\left(z_i^{kj}\ln\left(\frac{z_i^{kj}}{y_i^{kj}}\right)+y_i^{kj}-z_i^{kj}\right)=\hat{z}_i\ln\left(\frac{\hat{z}_i}{y_i}\right)+y_i-\hat{z}_i.$$

Como $z_i^{kj}\ln\left(\frac{z_i^{kj}}{y_i^{kj}}\right)+y_i^{kj}-z_i^{kj}\to 0$ entonces $\hat{z}_i\ln\left(\frac{\hat{z}_i}{y_i}\right)+y_i-\hat{z}_i=0$.
Luego,

$$\begin{aligned}
y_i\left(\frac{\hat{z}_i}{y_i}\ln\left(\frac{\hat{z}_i}{y_i}\right)+1-\frac{\hat{z}_i}{y_i}\right)=0 &\Rightarrow \frac{\hat{z}_i}{y_i}\ln\left(\frac{\hat{z}_i}{y_i}\right)-\frac{\hat{z}_i}{y_i}+1=0\\
&\Rightarrow \varphi\left(\frac{\hat{z}_i}{y_i}\right)=\varphi_1\left(\frac{\hat{z}_i}{y_i}\right)=0\\
&\Rightarrow \frac{\hat{z}_i}{y_i}=1, \qquad\qquad \text{(Por proposición 2.1(a))}\\
&\Rightarrow \hat{z}_i=y_i.
\end{aligned}$$

Por lo tanto, toda subsucesión convergente de $\{z^k\}$ converge a y. $(*)$
Supongamos que $\{z^k\}$ no es convergente.
Como $\{z^k\}\subset\mathbb{R}^n_+$ es acotada, entonces existen dos subsuceciones convergentes a dos puntos distintos. Lo cual contradice $(*)$. Por consiguiente, $\{z^k\}$ es convergente.
Así, por $(*)$ $\{z^k\}$ converge a y.

$\blacksquare$

Lema 2.4. *Para todo $x, y \in \mathbb{R}^n_{++}$ y $z \in \mathbb{R}^n_{+}$, se tiene que:*

(a) *Si $\varphi \in \Phi_2$, entonces $H(z,x) - H(z,y) \geq (\varphi''(1))^{-1}\sum_{i=1}^{n}(z_i - y_i)\varphi'\left(\frac{y_i}{x_i}\right)$, donde H está dada por (2.12).*

(b) *Si $\varphi(t) = t - \ln(t) - 1$, entonces $d(x,z) - d(y,z) \geq \sum_{i=1}^{n}(z_i - y_i)\varphi'\left(\frac{y_i}{x_i}\right)$, donde d está dada por (2.14).*

Demostración:

Parte **(a)**

$$
\begin{aligned}
H(z,x) - H(z,y) &= \sum_{i=1}^{n}\left[z_i \log\left(\frac{z_i}{x_i}\right) + x_i - z_i\right] - \sum_{i=1}^{n}\left[z_i \log\left(\frac{z_i}{y_i}\right) + y_i - z_i\right] \\
&= \sum_{i=1}^{n} z_i \log\left(\frac{z_i}{x_i}\right) + \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} z_i - \sum_{i=1}^{n} z_i \log\left(\frac{z_i}{y_i}\right) - \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} z_i \\
&= \sum_{i=1}^{n} z_i \left(\log\left(\frac{z_i}{x_i}\right) - \log\left(\frac{z_i}{y_i}\right)\right) + \sum_{i=1}^{n}(x_i - y_i) \\
&= \sum_{i=1}^{n} z_i \log\left(\frac{y_i}{x_i}\right) + \sum_{i=1}^{n}(x_i - y_i) \\
&= \sum_{i=1}^{n} z_i \log\left(\frac{y_i}{x_i}\right) + \sum_{i=1}^{n} y_i\left(\frac{x_i}{y_i} - 1\right) \\
&\geq [\varphi''(1)]^{-1}\sum_{i=1}^{n} z_i\varphi'\left(\frac{y_i}{x_i}\right) - [\varphi''(1)]^{-1}\sum_{i=1}^{n} y_i\varphi'\left(\frac{y_i}{x_i}\right) \\
&= [\varphi''(1)]^{-1}\sum_{i=1}^{n}(z_i - y_i)\varphi'\left(\frac{y_i}{x_i}\right).
\end{aligned}
$$

Parte **(b)**. Sea $\varphi(t) = t - \log t - 1$. Entonces, $\varphi''(1) = 1$.

$$
\begin{aligned}
d(x,z) - d(y,z) &= \sum_{i=1}^{n}\left[z_i \log\left(\frac{z_i}{x_i}\right) + x_i - z_i\right] - \sum_{i=1}^{n}\left[z_i \log\left(\frac{z_i}{y_i}\right) + y_i - z_i\right] \\
&= H(z,x) - H(z,y) \\
&\geq [\varphi''(1)]^{-1}\sum_{i=1}^{n}(z_i - y_i)\varphi'\left(\frac{y_i}{x_i}\right) \qquad \text{(por parte (a))} \\
&= \sum_{i=1}^{n}(z_i - y_i)\varphi'\left(\frac{y_i}{x_i}\right).
\end{aligned}
$$

■

Capítulo 3

Análisis de Convergencia.

§3.1. Un algoritmo punto proximal con $\varphi-$divergencia.

En esta sección analizaremos el algoritmo punto proximal con $\varphi - divergencia$ e introduciremos las hipótesis básica y determinaremos que el algoritmo está bien definido, el cual abreviaremos por DPM.

Algoritmo DPM:

Inicialización: Sea $x^0 \in \mathbb{R}^n$, $x^0 > 0$ y un criterio de parada

$k = 0$

mientras el criterio de parada no se cumpla

Si $\nabla f(x^k) = 0 \rightarrow parar$

$T_{k+1} = \{z \in \mathbb{R}^n_+ : z = \arg\min\{f(x) + \lambda_k d(x, x^k), x \in \mathbb{R}^n_+\}\}$

tomar $x^{k+1} \in T_{k+1}$

$k = k + 1$

fin (fin del mientras)

fin Algoritmo DPM

donde $d(x, y)$ está dado por (2.14).

Recordemos que:

$$d(x, y) = \sum_{i=1}^{n} y_i \varphi\left(\frac{x_i}{y_i}\right) = \sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{x_i}\right) + (x_i - y_i)\right]. \tag{3.1}$$

Así que,

$$(\nabla_1 d(x, y))_i = \varphi^{'}\left(\frac{x_i}{y_i}\right) = -\frac{y_i}{x_i} + 1. \tag{3.2}$$

donde ∇_1 representa la derivada respecto a la primera variable.

Vamos a considerar dos hipótesis básicas sobre f que son necesarias para probar los siguientes resultados:
Suposición 1: f es continuamente diferenciable, es decir, $f \in C^1$;
Suposición 2: (P) tiene un conjunto solución no vacío X_*.

Tal como se indica en el algoritmo, denotaremos:

$$T_{k+1} = \{z \in \mathbb{R}^n_+ : z = \arg\min f_k(x), x \in \mathbb{R}^n_+\} \quad \text{para todo } k \geq 0, \tag{3.3}$$

donde $f_k(x) := f(x) + \lambda_k d(x, x^k)$. Haremos $T_0 = \{x^0\}$.
Se establece así la buena definición del algoritmo de DPM a través de dos resultados. En la primera, se garantiza la existencia de x^k en cada iteración x^k y, además, que $x^k > 0$ para cada $k \in \mathbb{N}$, es decir, tenemos un algoritmo de punto interior. El segundo resultado demuestra que dos cualesquiera iterados son distintos, es decir, $x^{k_1} \neq x^{k_2}$ siempre que $k_1 \neq k_2$. Por lo tanto, el algoritmo DPM tiene la propiedad no cíclica.

Teorema 3.1. *Para todo $k \geq -1$, $T_{k+1} \bigcap \mathbb{R}^n_{++} \neq \emptyset$. Equivalentemente, existe $x^{k+1} \in T_{k+1}$ con $x^{k+1} > 0$, para todo $k \geq -1$.*

Demostración:
Procedemos por inducción. Sabemos que $x^0 \in T_0$ y $x^0 > 0$. Probemos que $x^k > 0$ implica que existe $x^{k+1} \in T_{k+1}$ con $x^{k+1} > 0$.
Sea $x^* \in X_* \neq \emptyset$. Entonces

$$f_k(x) = f(x) + \lambda_k d(x, x^k) \qquad \geq f(x^*) + \lambda_k d(x, x^k) \geq f(x^*), \quad \forall x \in \mathbb{R}^n_{++},$$

debido a que $d(x, x^k) \geq 0$ y $\lambda_k > 0$.
Así, f_k esta acotada inferiormente por $f(x^*)$.
Puesto que $x_i^k > 0$ y usando las partes (c) y (d) de la proposición (2.1), se tiene que:

$$x_i^k \varphi\left(\frac{x_i}{x_i^k}\right) \to +\infty \quad \text{cuando} \quad x_i^k \to 0^+ \qquad x_i^k \to +\infty.$$

De ese modo,

$$\lambda_k d(x, x^k) = \lambda_k \sum_{i=1}^{n} x_i^k \varphi\left(\frac{x_i}{x_i^k}\right) \to +\infty \quad \text{cuando} \quad x \to \partial\mathbb{R}^n_{++}. \tag{3.4}$$

Dado que $\lambda_k > 0$ y debido a la positividad de x_i^k y φ, cada una de las partes de la suma es no negativa. Ahora, dado que $X_* \neq \emptyset$, f es acotado inferiormente y (3.4) se cumple, entonces:

$$f_k(x) = f(x) + \lambda_k d(x, x^k) \to +\infty \quad \text{cuando} \quad x \to \partial\mathbb{R}^n_{++},$$

Puesto que f es continua, acotada inferiormente y sastiface que $f_k(x) \to +\infty$ cuando $x \to \partial\mathbb{R}^n_{++}$, esto garantiza que f_k alcanza su mínimo en un punto $\hat{x} > 0$. Por consiguiente, existe $x^{k+1} \in T_{k+1}$ con $x^{k+1} > 0$, para todo $k \geq -1$. ■

Observación 3.1. *$T_{k+1} \bigcap \partial\mathbb{R}^n_{++} = \emptyset$, para todo $k \geq -1$.*

Proposición 3.1. *Si $z \in T_{k+1}$, entonces $\nabla f(z) = -\lambda_k \nabla_1 d(z, x^k)$, para todo $k \geq 0$. En particular, la sucesión $\{x^k\}$ generado por el algoritmo DPM sastiface $\nabla f(x^{k+1}) = -\lambda_k \nabla_1 d(x^{k+1}, x^k)$, para todo $k \geq 0$.*

Demostración:

Sea $k \geq 0$. Las condiciones de optimalidad necesarias para minimizar $f_k(x) = f(x) + \lambda_k d(x, x^k)$ bajo $x \geq 0$ se expresan como:

$$x \geq 0, \nabla f_k(x) \geq 0 \quad y \quad x_i(\nabla f_k(x))_i = 0, \quad i = 1, ..., n.$$

Si $z \in T_{k+1}$ entonces, por Teorema 3.1, z es una solución óptima para f_k y $z > 0$. Así, se tiene $\nabla f_k(z) = 0$. Es decir, $\nabla f(z) + \lambda_k \nabla_1 d(z, x^k) = 0$. Por lo tanto, $\nabla f(z) = -\lambda_k \nabla_1 d(z, x^k)$. En particular, hemos tomado en el algoritmo $x^{k+1} \in T_{k+1}$, por lo que se tiene que $\nabla f(x^{k+1}) = -\lambda_k \nabla_1 d(x^{k+1}, x^k)$. ■

Lema 3.2. *$T_{k+1} \bigcap T_k = \emptyset$, para todo $k \geq 0$. De manera equivalente, la suseción $\{x^k\}$ generado por el algoritmo DPM sastiface $x^{k+1} \neq x^k$, para todo $k \geq 0$, es decir, dos iteraciones consecutivas son distintas.*

Demostración:

Sea $k \geq 0$. Supongamos que $\nabla f(x^k) \neq 0$, es decir, el algoritmo continúa. Tomemos $x^{k+1} \in T_{k+1}$ y supongamos que $x^{k+1} = x^k$. Entonces, por Proposición 3.1, se tiene

$$\nabla f(x^{k+1}) = -\lambda_k \nabla_1 d(x^{k+1}, x^k) = -\lambda_k \nabla_1 d(x^k, x^k) = 0.$$

De ahí, $\nabla f(x^k) = \nabla f(x^{k+1}) = 0$. Pero esto contradice la hipótesis de que $\nabla f(x^k) \neq 0$.

Por lo tanto, $x^{k+1} \neq x^k$, para todo $k \geq 0$. ■

Teorema 3.3. *$T_{k+1} \bigcap T_l = \emptyset$, para todo $k \geq 0$ y para todo l, $0 \leq l \leq k$, es decir, $x^{k+1} \neq x^l$, para todo $k \geq 0$ y para todo l, $0 \leq l \leq k$, es decir, la sucesión $\{x^k\}$ generada por el algoritmo DPM no es un ciclo.*

Demostración:
El caso $l = k \geq 0$ corresponde al Lema 3.2. Entonces, hacemos $k \geq 1$ y $l < k$.
Puesto que

$$x^{k+1} = \arg \min_{x \geq 0} f_k(x) = \arg \min_{x \geq 0} \{f(x) + \lambda_k d(x, x^k)\},$$

se tiene,

$$f(x^{k+1}) + \lambda_k d(x^{k+1}, x^k) \leq f(x^k) + \lambda_k d(x^k, x^k) = f(x^k). \tag{3.5}$$

Por otro lado, usando el Lema 3.2, $d(x^{k+1}, x^k) > 0$, para todo $k \geq 0$. Así, (3.5) nos proporciona

$$f(x^{k+1}) < f(x^k).$$

Por consiguiente, la sucesión $\{f(x^k)\}$ es estrictamente decreciente, es decir,

$$f(x^0) > f(x^1) > ... > f(x^k) > f(x^{k+1}) > ...$$

Supongamos que $x^{k+1} = x^l$ para algún $k \geq 1$ y $l < k$. Esto implica que $f(x^l) + \lambda_k d(x^l, x^k) \leq f(x^k)$. Dado que, $x^l = x^{k+1} \neq x^k$ y $\lambda_k > 0$, tenemos que $\lambda_k d(x^l, x^k) > 0$. Así que, $f(x^l) < f(x^k)$ y esto contradice el hecho de que $\{f(x^k)\}$ es estríctamente decreciente. Por tanto, $x^{k+1} \neq x^l$, para todo $k \geq 0$ y para todo l, $0 \leq l \leq k$. ■

§3.2. Análisis de Convergencia

En esta sección, vamos a demostrar que la sucesión $\{x^k\}$ generado por el algoritmo DPM es acotada y, además, que converge a un punto estacionario del problema (P). Por último, bajo la condición de $\lim_{k \to +\infty} \lambda_k = 0$, vamos a mostrar la convergencia con una solución del problema (P).

Lema 3.4. *Sea $\{x^k\}$ la suseción generada por el algoritmo DPM. Entonces,*

(a) *$0 < \lambda_k d(x^{k+1}, x^k) \leq f(x^k) - f(x^{k+1})$, para todo $k \geq 0$;*

(b) $\{f(x^k)\}$ *es una sucesión estrictamente decreciente y convergente;*

(c) $\sum_{k=0}^{\infty} \lambda_k d(x^{k+1}, x^k) < \infty$, *así,* $\lim\limits_{k\to+\infty} \lambda_k d(x^{k+1}, x^k) = 0$.

Demostración:

Parte **(a)**. Dado que $x^{k+1} \neq x^k$ y $\lambda_k > 0$, para todo $k \geq 0$, de ello se deduce que $\lambda_k d(x^{k+1}, x^k) > 0$ $\forall k \geq 0$. Para demostrar la segunda desigualdad, observemos que $x^{k+1} \in T_{k+1}$.

Luego, tenemos que

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

Así,

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

Parte **(b)**. Es consecuencia directa de la parte (a) y el hecho de que $X_* \neq \emptyset$.

Parte **(c)**. Por la parte (a), se tiene

$$\begin{aligned}
\sum_{k=0}^{K} \lambda_k d(x^{k+1}, x^k) &\leq \sum_{k=0}^{K} [f(x^k) - f(x^{k+1})] \\
&= f(x^0) - f(x^{K+1}) \\
&\leq f(x^0) - f(x^*) \\
&< +\infty,
\end{aligned}$$

donde $x^* \in X_*$.

Luego,

$$\sum_{k=0}^{\infty} \lambda_k d(x^{k+1}, x^k) = \lim_{K\to+\infty} \sum_{k=0}^{K} \lambda_k d(x^{k+1}, x^k) < +\infty.$$

■

Teorema 3.5. *Si f es cuasiconvexa y λ_k sastiface (3), la sucesión $\{x^k\}$ generada por el algoritmo DPM es acotada.*

Demostración:

Sea $x^* \in X_* \neq \emptyset$. De ahí, $f(x^*) \leq f(x^{k+1})$. Tomando $x = x^k$, $y = x^{k+1}$ y $z = x^*$ en la parte (b) del Lema 2.4, se obtiene

$$\begin{aligned} d(x^k, x^*) - d(x^{k+1}, x^*) &\geq \sum_{i=1}^{n} (x_i^* - x_i^{k+1}) \varphi' \left(\frac{x_i^{k+1}}{x_i^k} \right) \\ &= \left\langle \nabla_1 d(x^{k+1}, x^k), x^* - x^{k+1} \right\rangle \qquad \text{(por (3.2))} \\ &= \left\langle -\lambda_k^{-1} \nabla f(x^{k+1}), x^* - x^{k+1} \right\rangle \qquad \text{(por (3.1))} \\ &= \lambda_k^{-1} \left\langle \nabla f(x^{k+1}), x^{k+1} - x^* \right\rangle. \end{aligned} \tag{3.6}$$

Puesto que $\lambda_k > 0$, f es cuasiconvexa y $f(x^*) \leq f(x^{k+1})$, usando Teorema 2.2, se tiene:

$$\lambda_k^{-1} \left\langle \nabla f(x^{k+1}), x^{k+1} - x^* \right\rangle \geq 0. \tag{3.7}$$

De (3.6) y (3.7), $d(x^{k+1}, x^*) \leq d(x^k, x^*)$. De tal modo, la sucesión $\{d(x^k, x^*)\}$ es decreciente, es decir,

$$d(x^0, x^*) \geq d(x^1, x^*) \geq ... \geq d(x^k, x^*) \geq d(x^{k+1}, x^*) \geq ....$$

Por lo tanto,

$$\{x^k\} \subset \{x \in \mathbb{R}^n_{++} : d(x, x^*) \leq d(x^0, x^*)\}.$$

Ahora, usando la parte (a) del Lema 2.3, los conjuntos de nivel de $d(\cdot, x^*)$ son acotadas y entonces, $\{x^k\}$ es acotada. ■

Teorema 3.6. *Si f es cuasiconvexa y λ_k sastiface (3) , la sucesión $\{x^k\}$ generada por el algoritmo DPM converge a un punto estacionario de (P).*

Demostración:

Sea $\bar{x} \in \mathbb{R}^n_+$ un punto límite de $\{x^k\}$, el cual existe por el Teorema 3.5. De la parte (b) del Lema 3.4,$f(x^{k+1}) \leq f(x^k)$, para todo $k \in \mathbb{N}$, se procede, por la continuidad de f, que $f(\bar{x}) \leq f(x^k)$, para todo $k \in \mathbb{N}$. Ahora, usando el mismo argumento que el de la prueba anterior, se tiene que $\{d(x^k, \bar{x})\}$ es decreciente. Por otro lado, ya que $\bar{x}$ es un punto límite de $\{x^k\}$, existe una subsucesión $\{x^{pk}\}$ de $\{x^k\}$ tal que $\lim\limits_{k \to +\infty} x^{pk} = \bar{x}$. Por la parte (b) del Lema 2.3, tenemos que $\lim\limits_{k \to +\infty} d(x^{pk}, \bar{x}) = 0$. Por consiguiente, $\{d(x^k, \bar{x})\}$ es una sucesión no negativa y decreciente que posee una subsucesión convergente a 0. Por tanto, toda la sucesión debe converger a 0, es decir,

$$\lim_{k\to+\infty} d(x^k, \bar{x}) = 0. \tag{3.8}$$

Ahora bien, sea $\tilde{x}$ otro punto límite de $\{x^k\}$. Entonces, existe una subsucesión $\{x^{qk}\}$ de $\{x^k\}$ de tal manera que $\lim_{k\to+\infty} x^{qk} = \tilde{x}$. De (3.8), también tenemos $\lim_{k\to+\infty} d(x^{qk}, \bar{x}) = 0$. Tomando $y^k = x^{qk}$,$y = \tilde{x}$ y $z^k = \bar{x}$ en la parte (c) del Lema 2.3, tenemos entonces que $\bar{x} = \tilde{x}$. Por lo tanto, la sucesión $\{x^k\}$ tiene un único punto límite, es decir, la sucesión $\{x^k\}$ converge.

Vamos a mostrar, ahora, que $\bar{x}$ es un punto estacionario del problema (P), que es

$$\bar{x} \geq 0, \quad \nabla f(\bar{x}) \geq 0 \quad y \quad \bar{x}_i(\nabla f(\bar{x}))_i = 0, \quad i = 1, ..., n. \tag{3.9}$$

Sea

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

La primera condición en (3.9) es obvia, ya que $\lim_{k\to+\infty} x^k = \bar{x}$ y $x^k > 0$. Para mejorar las demás condiciones de (3.9), vamos a considerar dos casos.

Caso **(i)**: Si $i \in I(\bar{x})$, vamos probar que $(\nabla f(\bar{x}))_i \geq 0$.

Sea $i \in I(\bar{x})$. Supongamos que $(\nabla f(\bar{x}))_i < 0$. Por la diferenciabilidad continua de f, tenemos que $(\nabla f(x^{k+1}))_i \to (\nabla f(\bar{x}))_i < 0$. De ese modo, $(\nabla f(x^{k+1}))_i < 0$ para k suficientemente grande. De la Proposición 3.1 y la ecuación (3.2), se tiene

$$(\nabla f(x^{k+1}))_i = -\lambda_k(\nabla_1 d(x^{k+1}, x^k))_i = -\lambda_k \varphi' \left(\frac{x_i^{k+1}}{x_i^k} \right).$$

Dado que $\lambda_k > 0$, $\varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) > 0$ para k suficientemente grande. Ahora, por las propiedades de $\varphi(t)$ de la sección 2.2, obtenemos que $x_i^{k+1} > x_i^k$ para k suficientemente grande. Por otra parte, la sucesión completa $\{x_i^k\}$ converge a $\bar{x}_i$ y $\bar{x}_i = 0$. Lo cual contradice el hecho de que $x_i^{k+1} > x_i^k > 0$ para k suficientemente grande. En conclusión, $(\nabla f(\bar{x}))_i \geq 0$, para todo $i \in I(\bar{x})$.

Caso **(ii)**: Si $i \in J(\bar{x})$, probaremos que $(\nabla f(\bar{x}))_i = 0$.
Sea $i \in J(\bar{x})$. Dado que $\lim_{k\to+\infty} x_i^k = \bar{x}_i > 0$, se tiene $\lim_{k\to+\infty} \dfrac{x_i^{k+1}}{x_i^k} = 1$. Nuevamente, de las propiedades de $\varphi(t)$ en la sección 2.2, se obtiene $\varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) \to \varphi'(1) = 0$. Además, de la Proposición 3.1 y la ecuación (3.2), tenemos que

$$(\nabla f(x^{k+1}))_i = -\lambda_k \varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right).$$

Por ello, $(\nabla f(x^k))_i \to (\nabla f(\bar{x}))_i = 0$, para todo $i \in J(\bar{x})$.

De los casos (i) y (ii), concluimos que:

$$\nabla f(\bar{x}) \geq 0 \quad y \quad \bar{x}_i(\nabla f(\bar{x}))_i = 0, \quad i = 1, ..., n.$$

Por lo tanto, $\bar{x}$ es un punto estacionario del problema (P).

■

Teorema 3.7. *Si f es cuasiconvexa y los parámetros λ_k sastifaciendo (3) y (4), la sucesión $\{x^k\}$ generada por el algoritmo DPM converge a una solución de (P).*

Demostración:
De (3.3), tenemos que:

$$x^{k+1} \in \arg\min\{f(x) + \lambda_k d(x, x^k), x \in \mathbb{R}_+^n\}.$$

Entonces,

$$f(x^{k+1}) + \lambda_k d(x^{k+1}, x^k) \leq f(x) + \lambda_k d(x, x^k), \quad \forall x > 0.$$

Usando que $\lim_{k\to+\infty} \lambda_k = 0$ por (4), los conjuntos de nivel de $d(\cdot, x^k)$ son acotados por la parte (a) del Lema 2.3 y $\lim_{k\to+\infty} \lambda_k d(x^{k+1}, x^k) = 0$ por la parte (c) del Lema 3.4, tenemos, tomando $k \to +\infty$,

$$f(\bar{x}) \leq f(x), \quad \forall x > 0, \tag{3.10}$$

donde $\bar{x}$ es tal que $\lim_{k\to+\infty} x^k = \bar{x}$. Sea $x^* \in X_* \neq \emptyset$ y tomemos una sucesión $y^k \in \mathbb{R}_{++}^n$ de tal manera que $\lim_{k\to+\infty} y^k = x^*$. De (3.10), se obtiene

$$f(\bar{x}) \leq f(y^k), \quad \forall k > 0.$$

Tomando nuevamente $k \to +\infty$, tenemos

$$f(\bar{x}) \leq f(x^*).$$

Sin embargo, esto es posible únicamente si $\bar{x} \in X_*$.

■

Bibliografía

[1] F.G.M. Cunha, J.X. da Cruz Neto, P.R. Oliveira. (June 2006). *A proximal point algorithm with $\varphi-$Divergence to quasiconvex programming*

[2] Bazaraa, M.S.,Sherali, H.D. and Sheetty, C.M. (1993). *Nonlinear Programming: Theory and Algorithms, 2nd Edition*, John Wiley and Sons Inc., New York.

[3] Csiszár, I. (1967). *Information-type measures of difference of probability distributions and indirect observations, Studia Scientiarum Mathematicarum Hungarica.*

[4] Iusem, A.N., Svaiter, B.F. and Teboulle, M. (1994), *Entropy like proximal methods in convex programming, Mathematics of Operations Research.*

[5] Iusem, A.N. and Teboulle, M. (1993), *On the convergence rate of entropic proximal optimization algorithm, Computational and Applied Mathematics.*

[6] Iusem, A.N. and Teboulle, M. (1995), *Convergence rate analysis of nonquadratic proximal methods for convex and linear programming Mathematics of Operations Research.*

[7] Martinet, B. (1978), *Perturbation des méthodes dópimisation, Application, R.A.I.R.O.,Analyse Numérique.*

[8] Rockafellar, R. T. (1976), *Augmented Lagrangians and applications of proximal point algorithm in convex programming, Mathematics of Operation Research.*

[9] Rockafellar, R. T. (1976), *Monotone operators and the proximal point algorithm, SIAM Journal on Control and Optimization.*

[10] Teboulle, M. (1992), *Entropic proximal mappings with applications to nonlinear programming, Mathematics of Operation Research.*

[11] Teboulle, M. (1997), *Convergence of proximal-like algorithms, SIAM Journal on Optimization.*