

Método del Punto Proximal para Optimización

Minicurso para el IV Congreso Internacional de Optimización y Aplicaciones

Dr. Erik Alex Papa Quiroz

Universidad Nacional Mayor de San Marcos

Facultad de Ciencias Matemáticas

Setiembre, 2025

Método del Punto Proximal para Optimización

Erik Alex Papa Quiroz

Setiembre/2025

Universidad Nacional Mayor de San Marcos.

RESUMEN

En este curso presentamos el método del punto proximal para minimizar problemas de minimización no necesariamente diferenciables. Probamos que cuando la función objetivo es convexa y cuasi-convexa el método genera una sucesión de puntos que converge a un punto solución (en el caso convexo) y a un punto candidato a solución (en el caso cuasi-convexo). Por otro lado, cuando la función tiene la propiedad de la desigualdad de Lojasiewicz probamos que toda sucesión acotada converge a un punto crítico generalizado del problema.

Palabras Claves: método del punto proximal, optimización convexa, optimización cuasi-convexa, desigualdad de Lojasiewicz.

Capítulo 1

Preliminares

En este capítulo presentamos las notaciones utilizadas e haremos un resumen de los resultados básicos necesarios y útiles para el buen entendimiento de este trabajo.

1.1 Notaciones

A lo largo de esta monografía, adoptaremos las siguientes notaciones:

$\mathbb{R}^n = \{x = (x_1, x_2, \dots, x_n) / x_i \in \mathbb{R}, \forall i = 1, \dots, n\}$: espacio Euclidiano.

$\mathbb{R}_+$: el conjunto de los números reales no negativos.

$\mathbb{R}_{++}$: el conjunto de los números reales positivos.

Dados $x, y \in \mathbb{R}^n$, $x^T y = \sum_{i=1}^n x_i y_i$: el producto interno Euclidiano en $\mathbb{R}^n$.

Dado $x \in \mathbb{R}^n$, $\|x\| := \sqrt{x^T x}$: la norma euclidiana.

$\text{dom} f$: dominio de la función f .

$\text{epi} f$: epigrafe de f .

$(0, 1) := \{z \in \mathbb{R} : 0 < z < 1\}$

$\hat{\partial} f(x)$: el subdiferencial de Fréchet de f en x .

∇f : el gradiente de f .

$\partial f(x)$: la subdiferencial en el limite de la función f .

$\max\{f(x), f(y)\}$: el valor máximo entre $f(x)$ y $f(y)$.

$B(\bar{x}, \epsilon) = \{x \in \mathbb{R}^n : \|x - \bar{x}\| \leq \epsilon\}$: la bola Euclidiana de centro $\bar{x}$ y radio ϵ .

1.2 Funciones Convexas y Cuasi-convexas

En esta sección damos algunos resultados de funciones convexas y cuasi-convexas.

Dada una función de valores extendidos $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, el dominio de f , denotado por $\text{dom} f$ es definido por

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

y su epigrafo por

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

f es llamada una función propia si $\text{dom} f \neq \emptyset$ y $\forall x \in \text{dom} f$ tenemos $f(x) > -\infty$. Decimos que f es una función semicontinua inferior si $\text{epi} f$ es un subconjunto cerrado de $\mathbb{R}^n \times \mathbb{R}$.

Definición 1.1 Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia. f es convexa en $\mathbb{R}^n$ si para todo $x, y \in \mathbb{R}^n$ y todo $t \in [0, 1]$, se cumple:

$$f(tx + (1-t)y) \leq tf(y) + (1-t)f(x),$$

Cuando la desigualdad es estricta, para $x \neq y$ y $t \in (0, 1)$, f es llamada estrictamente convexa.

Definición 1.2 Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función real de valores extendidos. f es llamada cuasi-convexa en $\mathbb{R}^n$ si para todo $x, y \in \mathbb{R}^n$ y todo $t \in [0, 1]$, se cumple que

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

Teorema 1.1 Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia con $\text{dom} f$ convexo en $\mathbb{R}^n$. f es cuasi-convexa en $\mathbb{R}^n$ si, y solamente si, para cada $c \in \mathbb{R}$ el conjunto de nivel inferior de f , $\Gamma_f(c) = \{x \in \mathbb{R}^n : f(x) \leq c\}$, es convexo.

Demostración. Supongamos que f es cuasi-convexa y sea $c \in \mathbb{R}$. Probaremos que $\Gamma_f(c)$ es convexo. Sea $x, y \in \Gamma_f(c)$, entonces $f(x) \leq c$, $f(y) \leq c$ y por tanto $\max\{f(x), f(y)\} \leq c$. Usando la cuasi-convexidad de f y el resultado previo, tenemos que

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

De la desigualdad anterior obtenemos que $\Gamma_f(c)$ es convexo.

Recíprocamente, supongamos que $\Gamma_f(c)$ es convexo para cada $c \in \mathbb{R}$. Probaremos que f es cuasi-convexa. Sea $x, y \in \text{dom} f$ y $t \in [0, 1]$. Definamos $c := \max\{f(x), f(y)\}$. De aquí tenemos inmediatamente que $x, y \in \Gamma_f(c)$ y como $\Gamma_f(c)$ es convexo, esto implica que

$$tx + (1 - t)y \in \Gamma_f(c).$$

Finalmente, por la definición de $\Gamma_f(c)$ obtenemos nuestro resultado. ■

Teorema 1.2 *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable y cuasi-convexa y sea $x, y \in \mathbb{R}^n$. Si $f(x) \leq f(y)$ entonces*

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

Demostración. Sea $t \in (0, 1)$. Por la aproximación de primer orden de una función diferenciable tenemos

$$f(y + t(x - y)) = f(y) + t\langle \nabla f(y), x - y \rangle + o(t),$$

donde

$$\lim_{t \rightarrow 0} \frac{o(t)}{t} = 0.$$

Como f es cuasi-convexa, esto implica que

$$f(y) + t\langle \nabla f(y), x - y \rangle + o(t) \leq \max\{f(x), f(y)\}.$$

Usando la hipótesis de que $f(x) \leq f(y)$ y dividiendo por t obtenemos

$$\langle \nabla f(y), x - y \rangle + \frac{o(t)}{t} \leq 0.$$

Tomando límite cuando t converge a cero, obtenemos el resultado. ■

El recíproco del teorema anterior es también válido, ver por ejemplo Teorema 3.5.4 de Bazaraa, Sherali y Shetty [7], pero aquí no lo desarrollaremos pues no será esencial en nuestro estudio.

Definición 1.3 *Una función diferenciable $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es pseudoconvexa si, para todo par de puntos distintos x, y tenemos*

$$\langle \nabla f(x), y - x \rangle \geq 0, \quad \text{entonces} \quad f(y) \geq f(x).$$

Teorema 1.3 Sea C un conjunto convexo cerrado no vacío en el espacio Euclidiano $\mathbb{R}^n$. Tomemos $y \in \mathbb{R}^n$ arbitrario, entonces existe una única proyección $z = P_C y$. Además, la siguiente desigualdad es satisfecha

$$\langle y - z, x - z \rangle \leq 0, \quad (1.1)$$

para todo $x \in C$.

Demostración. Ver Teorema 2.4.1 de Bazaraa Serali y Shety, [7]. ■

1.3 Teoría Subdiferencial

En esta sección recordamos algunas definiciones y resultados de subdiferencial de Fréchet y de subdiferencial en el límite. Para mayores detalles ver los capítulos 8 y 10 de Rockafellar y Wets [23].

Definición 1.4 Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia semicontinua inferior. Para cada $x \in \text{dom} f$, el subdiferencial de Fréchet de f en x , denotado por $\hat{\partial} f(x)$, es el conjunto de puntos $s \in \mathbb{R}^n$ tal que

$$\liminf_{y \neq x, y \rightarrow x} \frac{1}{\|x - y\|} [f(y) - f(x) - \langle s, y - x \rangle] \geq 0.$$

Si $x \notin \text{dom} f$ entonces $\hat{\partial} f(x) = \emptyset$.

Observemos que la definición anterior es equivalente a

$$f(y) \geq f(x) + \langle s, y - x \rangle + o(\|x - y\|),$$

donde

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

Puede ser demostrado que $\hat{\partial} f(x)$ es un conjunto cerrado y convexo. Sin embargo, una desventaja de trabajar con el subdiferencial de Fréchet es que ella no cumple algunas propiedades en el límite que necesitaremos, por eso introducimos la siguiente definición.

Definición 1.5 Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia semicontinua inferior. El subdiferencial en el límite de f en $x \in \mathbb{R}^n$, denotado por $\partial f(x)$, es definido como sigue

$$\partial f(x) := \{s \in \mathbb{R}^n : \exists x^n \rightarrow x, f(x^n) \rightarrow f(x), s^n \in \hat{\partial} f(x^n) \rightarrow s\}$$

Puede ser probado, ver el capítulo 8 de Rockafellar y Wets [23], que si f es una función propia y semicontinua inferior, los dominios de $\hat{\partial} f$ y ∂f son densos en el dominio de f .

Proposición 1.1 Las siguientes propiedades son satisfechas:

- a. $\hat{\partial} f(x) \subset \partial f(x)$, para todo $x \in \mathbb{R}^n$;
- b. Si f es diferenciable en $\bar{x}$ entonces $\hat{\partial} f(\bar{x}) = \{\nabla f(\bar{x})\}$, donde ∇f es el gradiente de f , así $\nabla f(\bar{x}) \in \partial f(\bar{x})$;
- c. Si f es diferenciable en una vecindad de x , entonces $\hat{\partial} f(x) = \partial f(x) = \{\nabla f(\bar{x})\}$;
- d. Si $g = f + h$ con f finito en $\bar{x}$ y h es diferenciable en una vecindad de $\bar{x}$ entonces

$$\hat{\partial} g(\bar{x}) = \hat{\partial} f(\bar{x}) + \nabla h(\bar{x}),$$

$$\partial g(\bar{x}) = \partial f(\bar{x}) + \nabla h(\bar{x}).$$

Demostración.

- a. Es inmediato a partir de la Definición 1.4 y Definición 1.5.
- b. Como f es diferenciable entonces:

$$f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + o(\|x - \bar{x}\|).$$

Sea $s \in \hat{\partial} f(\bar{x})$, de la Definición 1.4 tenemos:

$$f(x) \geq f(\bar{x}) + \langle s, x - \bar{x} \rangle + o(\|x - \bar{x}\|).$$

Usando la aproximación de primer orden en f en $\bar{x}$ y de la desigualdad anterior obtenemos:

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq \langle s, x - \bar{x} \rangle + o(\|x - \bar{x}\|).$$

Tomando $x = \bar{x} + \lambda(v - \nabla f(\bar{x}))$ con λ suficientemente pequeño, dá $v = \nabla f(\bar{x})$, como queríamos.

c. Sea $s \in \partial f(\bar{x})$, entonces

$$\exists x^n \rightarrow x, f(x^n) \rightarrow f(x), s^n \in \widehat{\partial} f(x^n) \rightarrow s.$$

De b. tenemos que $s^n = \nabla f(x^n)$, para n suficientemente grande. De la continuidad de ∇f y unicidad del límite obtenemos finalmente que $s = \nabla f(\bar{x})$.

d. La inclusión $\subset$ es inmediata. Para obtener la segunda inclusión $\supset$ aplicamos esta regla a la función $f = g + (-h)$. ■

1.4 Teoría de Fejér-convergencia

En esta sección damos un resultado fundamental en la teoría de Fejér-convergencia en espacios Euclidianos, el cual será utilizado para facilitar la prueba de convergencia global de los métodos presentados.

Definición 1.6 Una sucesión $\{y^k\}$, $k \geq 0$, de $\mathbb{R}^n$ es llamado cuasi-Fejér convergente al conjunto $U \subset X$, si para todo $u \in U$ existe una sucesión $\{\epsilon_k\} \subseteq \mathbb{R}$ tal que $\epsilon_k \geq 0$, $\sum_{k=0}^{+\infty} \epsilon_k < +\infty$ y

$$\|y^{k+1} - u\|^2 \leq \|y^k - u\|^2 + \epsilon_k.$$

Si $\epsilon_k = 0$, para todo k , entonces decimos que $\{y^k\}$ es Fejér convergente al conjunto U .

Teorema 1.4 Si $\{y^k\}$ es cuasi-Fejér convergente a un subconjunto no vacío $U \subseteq \mathbb{R}^n$, entonces $\{y^k\}$ es limitado. Si, además, un punto de acumulación $\bar{y}$ de $\{y^k\}$ pertenece a U , entonces $\{y^k\}$ converge y

$$\lim_{k \rightarrow +\infty} y^k = \bar{y}.$$

Demostración. Ejercicio . ■

Capítulo 2

Método del Punto Proximal

Consideremos el problema de optimización:

$$\min\{f(x) : x \in \mathbb{R}^n\}, \quad (2.1)$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ es una función propia semicontinua inferior. El método del punto proximal, para resolver el problema (2.1) genera una sucesión de puntos $\{x^k\}$ definidos por:

$$x^0 \in \mathbb{R}^n, \quad (2.2)$$

$$x^k \in \arg \min\{f(x) + (\lambda_k/2)\|x - x^{k-1}\|^2 : x \in \mathbb{R}^n\}. \quad (2.3)$$

donde λ_k es algún parámetro positivo y la notación $\arg \min$ significa que x^k es un mínimo de la función $f(x) + (\lambda_k/2)\|x - x^{k-1}\|^2$.

Este algoritmo fue introducido por Martinet [16] y fue muy estudiado por Rockafellar, [22] en los años setenta del siglo anterior para resolver problemas de optimización y más generalmente para resolver problemas de encontrar ceros de operadores monótonos maximales.

Los resultados de convergencia del método para el caso convexo son muy bien conocidos y puede ser probado la convergencia para el ínfimo de la función f y convergencia a un punto de mínimo si el mínimo existe, ver por ejemplo la Sección 2 de Iusem [12] o también el artículo de Guler [10]. En la actualidad se están realizando investigaciones para extender las aplicaciones del método para funciones no necesariamente convexas,

motivados por las múltiples aplicaciones en Economía, así como también, para acelerar la convergencia de lineal a superlineal.

En este capítulo, presentamos los resultados de convergencia del método para los casos donde la función objetivo f es convexa, cuasi-convexa y analítica. Para el caso convexo, presentamos los resultados de convergencia de Güler [10]. Entretanto, para el caso cuasi-convexo presentamos nuestros resultados obtenidos en Papa Quiroz y Oliveira [20]. Finalmente para el caso analítico, tomamos los resultados de Attouch y Bolte [2].

2.1 Una Motivación en Teoría de la Decisión

Este ejemplo fue tomado del artículo de Souza et al. [24]. Sea el estado $X = \mathbb{R}_+^n$, interpretado económicamente como el conjunto de alternativas, ejecuciones, acciones, decisiones, etc. El término agente será referido a una persona, empresa o institución encargada de la toma de decisiones.

Un agente tiene una función objetivo

$$g : X \rightarrow \mathbb{R},$$

que representa la ganancia por unidad de tiempo (utilidad, pago, ingreso, beneficio). Consideremos $x \in X$ algún estado inicial dado y supongamos que el supremo (meta propuesta a lograr) de este objetivo es finito, esto es,

$$\bar{g} = \sup \{g(x), x \in X\} < +\infty.$$

Supongamos que el agente comienza de algún estado inicial $x_0 = x \in X$, conoce el supremo de g y el valor inicial $g(x_0)$. Así, él conoce cuan lejos está del supremo. Esta diferencia o llamada también salto es definida por:

$$n(x) = \bar{g} - g(x).$$

La manera más natural de que el agente alcance su meta es encontrar el objetivo de manera progresiva dividiéndolo por periodos que detallamos a continuación.

Fase Exploratoria: El agente explora en un entorno de x , denotado por: $E(x, r(x)) \subset X$, durante algún tiempo $\tau > 0$, para descubrir $y \in E(x, r(x)) \subset X$.

Fase Desplazamiento: El agente se desplaza de $x \in X$ al nuevo estado $y \in X$ en un tiempo $\delta > 0$.

Fase Explotación: El agente debe beneficiarse al obtener un nuevo estado mejorado. Él escoge explotar su beneficio en un tiempo $\mu > 0$.

Es así que del estado $x \in X$ descubrimos un nuevo estado $y \in X$ mejorado, tal que $g(y) > g(x)$, de modo que $n(y) < n(x)$.

En cada periodo $k \in \mathbb{N}$, se considera la función ganancia proximal:

$$P_k(x^k, y) = g(y) - \lambda_k C(x^k, y, \delta)$$

donde: $x^k, y \in E(x^k, r(x^k))$,

$$\lambda_k = \frac{1}{\mu_k},$$

$C(x^k, y, \delta)$ es el costo de desplazamiento satisfaciendo $C(x^k, x^k, 0) = 0$.

Asumiremos por simplicidad que $E(x^k, r(x^k)) = X$, $\forall k \in \mathbb{N}$. Así, el agente tratará de maximizar su ganancia proximal resolviendo para cada k :

$$x^{k+1} \in \operatorname{argmax}\{P_k(x_k, y) : y \in X\},$$

el cual es equivalente a:

$$x^{k+1} \in \operatorname{argmin}\{n(y) + \lambda_k C(x^k, y, \delta) : y \in X\}.$$

2.2 Hipótesis y Primeros Resultados

Asumimos inicialmente las siguientes hipótesis:

Hipótesis A. $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$ es limitada inferiormente.

Hipótesis B. $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$ es continua sobre $\operatorname{dom} f$.

Comenzaremos probando la buena definición de la sucesión $\{x^k\}$ generada por el método del punto proximal.

Teorema 2.1 *Asumamos que las hipótesis A y B son satisfechas. Entonces, la sucesión $\{x^k\}$, generada por el método del punto proximal, es bien definida.*

Demostración. Probaremos inicialmente la existencia de los iterados y posteriormente la propiedad de no ciclaje. Procedemos por inducción. Para $k = 0$ se cumple debido a (2.2). Asumese ahora que x^k es bien definida. Sea α la cota inferior de f sobre $\mathbb{R}^n$ entonces

$$f(x) + (\lambda_k/2)\|x - x^k\|^2 \geq \alpha + (\lambda_k/2)\|x - x^k\|^2,$$

para todo $x \in \mathbb{R}^n$. Siguese que los conjuntos de nivel de la función $f(\cdot) + (\lambda_k/2)\|\cdot - x^k\|^2$ son limitadas. También, esta función es continua sobre $\text{dom} f$, debido a la hipótesis sobre f . Por eso, los conjuntos de nivel de $f(\cdot) + \lambda\|\cdot - x^k\|^2$ son cerrados, de aqui compactos. De los argumentos de continuidad y compacidad tenemos que $f(\cdot) + (\lambda_k/2)\|\cdot - x^k\|^2$ tiene un mínimo global el cual puede ser no único debido a la no convexidad de f . En tal caso, x^{k+1} puede ser arbitrariamente elegido entre el conjunto de todos los minimizadores de $f(\cdot) + (\lambda_k/2)\|\cdot - x^k\|^2$. ■

Proposición 2.1 *Con las hipótesis A y B, se obtienen los siguientes resultados*

- a. $\{f(x^k)\}$ es no creciente.
- b. $\sum_{k=1}^{\infty} \lambda_k \|x^k - x^{k-1}\|^2 < +\infty$

Prueba.

- a. De (2.3), la condición de mínimo implica que, para todo k ,

$$f(x^k) + (\lambda_k/2)\|x^k - x^{k-1}\|^2 \leq f(x^{k-1}). \quad (2.4)$$

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

- b. De la desigualdad (2.4) y la hipótesis A obtenemos

$$\sum_{k=1}^n (\lambda_k/2)\|x^k - x^{k-1}\|^2 \leq f(x^0) - \alpha,$$

donde α es el limite inferior de f . Tomando $n \rightarrow \infty$ obtenemos el resultado. ■

2.3 Caso Convexo

En esta sección asumimos que f es convexa el cual implica que $\text{dom} f$ convexo y que los parámetros λ_k satisfacen

$$\sum_{k=1}^{+\infty} (1/\lambda_k) = +\infty.$$

Por la convexidad de f obtenemos que $\{x^k\}$ cumple

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

Por eso, el punto x^{k+1} satisfaze

$$\lambda_k(x^k - x^{k+1}) \in \partial f(x^{k+1}) \quad (2.5)$$

Lemma 2.1 *Para todo $x \in \mathbb{R}^n$ se cumple:*

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 - \|x^k - x^{k+1}\|^2 + (2/\lambda_k)(f(x) - f(x^{k+1})). \quad (2.6)$$

Prueba. Para todo $x \in \mathbb{R}^n$ tenemos

$$\begin{aligned} \|x^k - x\|^2 &= \langle (x^k - x^{k+1}) + (x^{k+1} - x), (x^k - x^{k+1}) + (x^{k+1} - x) \rangle \\ &= \|x^k - x^{k+1}\|^2 + \|x^{k+1} - x\|^2 - (2/\lambda_k) \langle \lambda_k(x^k - x^{k+1}), x - x^{k+1} \rangle. \end{aligned}$$

Usando (2.5) tenemos

$$\|x^k - x\|^2 \geq \|x^k - x^{k+1}\|^2 + \|x^{k+1} - x\|^2 + (2/\lambda_k)(f(x^{k+1}) - f(x)).$$

De esta desigualdad, obtenemos el resultado deseado. ■

Lemma 2.2 *Sea $\sigma_n = \sum_{k=1}^n (1/\lambda_k)$, entonces*

$$2\sigma_n(f(x^n) - f(x)) \leq \|x - x^0\|^2 - \|x - x^n\|^2 - \sum_{k=1}^n \lambda_k \sigma_k \|x^k - x^{k-1}\|^2.$$

Prueba. De (2.3) tenemos

$$f(x^k) + (\lambda_k/2)\|x^k - x^{k-1}\|^2 \leq f(x) + (\lambda_k/2)\|x - x^{k-1}\|^2,$$

para todo $x \in \mathbb{R}^n$. Tomando $x = x^{k-1}$ tenemos

$$(\lambda_k/2)\|x^k - x^{k-1}\|^2 \leq f(x^{k-1}) - f(x^k).$$

Definamos $\sigma_k = (1/\lambda_k) + \sigma_{k-1}$ con $\sigma_0 = 0$. De la última desigualdad tenemos

$$\begin{aligned} (\lambda_k/2)\sigma_{k-1}\|x^k - x^{k-1}\|^2 &\leq \sigma_{k-1}(f(x^{k-1}) - f(x^k)) \\ &= \sigma_{k-1}f(x^{k-1}) - (\sigma_k - (1/\lambda_k))f(x^k) \\ &= \sigma_{k-1}f(x^{k-1}) - \sigma_k f(x^k) + (1/\lambda_k)f(x^k). \end{aligned}$$

Tomando sumatoria sobre k , de 1 a n y multiplicando por 2 tenemos

$$\sum_{k=1}^n \lambda_k \sigma_{k-1} \|x^k - x^{k-1}\|^2 \leq -2\sigma_n f(x^n) + \sum_{k=1}^n (2/\lambda_k) f(x^k) \quad (2.7)$$

Por otro lado, del Lema anterior tenemos que

$$(2/\lambda_k)(f(x^k) - f(x)) \leq \|x - x^{k-1}\|^2 - \|x^k - x^{k-1}\|^2 - \|x - x^k\|^2.$$

Tomando sumatoria de 1 a n ,

$$\sum_{k=1}^n (2/\lambda_k)(f(x^k) - f(x)) \leq \sum_{k=1}^n (\|x - x^{k-1}\|^2 - \|x^k - x^{k-1}\|^2 - \|x - x^k\|^2).$$

Esto implica que

$$\sum_{k=1}^n (2/\lambda_k) f(x^k) \leq 2\sigma_n f(x) + \|x - x^0\|^2 - \|x - x^n\|^2 - \sum_{k=1}^n \|x^k - x^{k-1}\|^2.$$

La desigualdad anterior y (2.7) dá

$$\sum_{k=1}^n \lambda_k \sigma_{k-1} \|x^k - x^{k-1}\|^2 + 2\sigma_n f(x^n) \leq 2\sigma_n f(x) + \|x - x^0\|^2 - \|x - x^n\|^2 - \sum_{k=1}^n \|x^k - x^{k-1}\|^2.$$

Así,

$$2\sigma_n(f(x^n) - f(x)) \leq \|x - x^0\|^2 - \|x - x^n\|^2 - \sum_{k=1}^n (1 + \lambda_k \sigma_{k-1}) \|x^k - x^{k-1}\|^2.$$

Como $1 + \lambda_k \sigma_{k-1} = \lambda_k \sigma_k$ entonces

$$2\sigma_n(f(x^n) - f(x)) \leq \|x - x^0\|^2 - \|x - x^n\|^2 - \sum_{k=1}^n \lambda_k \sigma_k \|x^k - x^{k-1}\|^2. \quad \blacksquare$$

Teorema 2.2 Sea $\{x^k\}$ la sucesión generada por el método del punto proximal. Para todo $x \in R^n$, tenemos

$$f(x^n) - f(x) \leq \frac{\|x - x^0\|^2}{2\sigma_n}, \quad (2.8)$$

donde $\sigma_n = \sum_{k=1}^n (1/\lambda_k)$.

Si $\sigma_n \rightarrow +\infty$ entonces

$$f(x^k) \rightarrow f^* := \inf f(x).$$

Ademas, si $X^* \neq \emptyset$, entonces $\{x^k\}$ converge para un minimizador de f y satisface

$$f(x^n) - f^* \leq \frac{\|x - x^0\|^2}{2\sigma_n}, \quad \forall x^* \in X^*$$

Demostración. Del Lema anterior, inmediatamente obtenemos (2.8).

Por definición de f^* , existe $v \in \mathbb{R}^n$ tal que para todo $\epsilon > 0$

$$f(v) < f^* + \epsilon.$$

Tomando $x = v$ en (2.8) tenemos

$$f(x^n) - f^* - \epsilon \leq f(x^n) - f(v) \leq \frac{\|v - x^0\|^2}{2\sigma_n}.$$

De aquí,

$$f(x^n) - f^* \leq \frac{\|v - x^0\|^2}{2\sigma_n} + \epsilon.$$

Tomando $n \rightarrow +\infty$ obtenemos

$$\lim_{n \rightarrow +\infty} f(x^n) - f^* \leq \epsilon.$$

Desde que ϵ es arbitrario tenemos

$$\lim_{n \rightarrow +\infty} f(x^n) = f^*.$$

Ahora, si el conjunto $X^* \neq \emptyset$, sea $\bar{x} \in X^*$ entonces

$$f(\bar{x}) \leq f(x^k), \forall k.$$

Sustituyendo $\bar{x}$ en (2.6) y usando la última desigualdad tenemos

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

Esto implica que

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

De aquí $\{x^k\}$ es Fejér convergente a X^* . Del Teorema 1.4, $\{x^k\}$ es limitada y así existe una subsucesión $\{x^{k_j}\}$ convergente. Sea $x^* \in \mathbb{R}^n$ tal que

$$\lim_{j \rightarrow +\infty} x^{k_j} = x^*.$$

Por continuidad de f tenemos

$$\lim_{j \rightarrow +\infty} f(x^{k_j}) = f(x^*).$$

Como $\{f(x^k)\}$ es una sucesión no creciente y limitada inferiormente entonces $\{f(x^k)\}$ es convergente. Además, como ella tiene una subsucesión convergente, tenemos que toda la sucesión converge para $f(x^*)$.

$$\lim_{k \rightarrow +\infty} f(x^k) = f(x^*).$$

De aquí, obtenemos que $x^* \in X^*$. Finalmente usando de nuevo el Teorema 1.4 tenemos

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

2.4 Caso Cuasi-convexo

En esta sección probaremos la convergencia de la sucesión $\{x^k\}$ a un punto crítico de f cuando esta función es cuasi-convexa. Para ello usaremos los parametros λ_k satisfaciendo:

$$0 < \lambda_k < \bar{\lambda}, \quad (2.9)$$

donde $\bar{\lambda} > 0$. Es necesario también sustituir las hipótesis A y B por las siguientes:

Hipótesis A' . El conjunto de soluciones óptimas del problema (2.1), denotada por X^* , es no vacío.

Hipótesis B' . $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ es cuasi-convexa y continua sobre $\text{dom} f$.

Ahora, definamos el siguiente conjunto

$$U := \{x \in \text{dom} f : f(x) \leq \inf_k f(x^k)\}.$$

De la hipótesis A' and B' este conjunto es no vacío cerrado y convexo (ver el Teorema 1.1 para la prueba de la convexidad del conjunto U).

Teorema 2.3 *Con las hipótesis A' and B' , la sucesión $\{x^k\}$, generada por el método del punto proximal, es Fejér convergente a U .*

Demostración. Desde que x^k satisface (2.3) tenemos

$$f(x^k) + (\lambda_k/2)\|x^k - x^{k-1}\|^2 \leq f(x) + (\lambda_k/2)\|x - x^{k-1}\|^2, \forall x \in \mathbb{R}^n. \quad (2.10)$$

De aquí, $\forall x \in U$ se cumple que

$$\|x^k - x^{k-1}\|^2 \leq \|x - x^{k-1}\|^2.$$

Portanto, x^k es la única proyección de x^{k-1} sobre U y del Teorema 1.3 tenemos que

$$\langle x - x^k, x^{k-1} - x^k \rangle \leq 0. \quad (2.11)$$

Por otro lado, tomando $x \in U$ tenemos

$$\begin{aligned} \|x - x^{k-1}\|^2 &= \langle (x - x^k) + (x^k - x^{k-1}), (x - x^k) + (x^k - x^{k-1}) \rangle \\ &= \|x - x^k\|^2 + \|x^k - x^{k-1}\|^2 - 2\langle x - x^k, x^{k-1} - x^k \rangle \end{aligned}$$

Usando (2.11) obtenemos que

$$0 \leq \|x^k - x^{k-1}\|^2 \leq \|x - x^{k-1}\|^2 - \|x - x^k\|^2 \quad (2.12)$$

para todo $x \in U$. Así

$$\|x - x^k\|^2 \leq \|x - x^{k-1}\|^2. \quad (2.13)$$

Por lo tanto, $\{x^k\}$ es Fejér convergente a U . ■

Proposición 2.2 *Con las hipótesis A' y B' , los siguientes enunciados son verdaderos*

a. *Para todo $x \in U$ la sucesión $\{\|x - x^k\|\}$ es no creciente y convergente;*

b. $\lim_{k \rightarrow +\infty} \|x^k - x^{k-1}\| = 0$;

Demostración.

a. De (2.13), $\{\|x - x^k\|\}$ es una sucesión no creciente y limitada inferiormente, de aquí ella es convergente.

b. Tomando limite cuando k va para el infinito en (2.12) y usando el resultado anterior obtenemos $\lim_{k \rightarrow +\infty} \|x^k - x^{k-1}\| = 0$. ■

Teorema 2.4 *Supongamos que las hipótesis A' y B' son satisfechas. Entonces la sucesión $\{x^k\}$ converge. Además, ella converge a un punto crítico de f .*

Demostración. De la Proposición anterior, $\{x^k\}$ es Fejér convergente a U , así $\{x^k\}$ es limitado (ver el Theorema 1.4). De aquí, existe $\bar{x}$ y una subsucesión $\{x^{k_j}\}$ de $\{x^k\}$ que converge a $\bar{x}$. De la continuidad de f obtenemos

$$\lim_{j \rightarrow +\infty} f(x^{k_j}) = f(\bar{x}).$$

Como $\{f(x^k)\}$ es una sucesión no creciente con una subsucesión que converge a $f(\bar{x})$, toda la sucesión converge a $f(\bar{x})$ y por lo tanto

$$f(\bar{x}) \leq f(x^k), \forall k \in \mathbb{N}.$$

Esto implica que $\bar{x} \in U$. Ahora, del Teorema 1.4 concluimos que $\{x^k\}$ converge a $\bar{x}$. Finalmente, probaremos que $0 \in \partial f(\bar{x})$. De (2.3) obtenemos

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

De la Proposición 1.1, **d**, y la regularidad de $(\lambda_k/2)\|\cdot - x^{k-1}\|^2$, esto implica que

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

Así, existe $g^k \in \widehat{\partial}f(x^k)$ tal que

$$\lambda_k(x^k - x^{k-1}) = -g^k.$$

Tomando $\|\cdot\|$ en la igualdad anterior y usando (2.9) obtenemos que

$$\|g^k\| \leq (\bar{\lambda})\|x^k - x^{k-1}\|.$$

Usando la Proposición 2.2, **b**, tenemos que

$$\lim_{k \rightarrow \infty} g^k = 0.$$

Portanto, existen sucesiones $\{x^k\}$, $\{f(x^k)\}$ y $\{g^k\}$ con $g^k \in \widehat{\partial}f(x^k)$ tal que $\lim_{k \rightarrow +\infty} x^k = \bar{x}$, $\lim_{k \rightarrow \infty} f(x^k) = f(\bar{x})$ (de la continuidad de f) y $\lim_{k \rightarrow \infty} g^k = 0$. De la Definición 1.5 siguese que $0 \in \partial f(\bar{x})$. ■

Observación 2.1 *Observemos que el resultado del teorema anterior es nuevo en la teoría del método del punto proximal para funciones cuasi-convexas.*

Como inmediatos casos particulares del Teorema precedente obtenemos los siguientes

Corolário 2.1 *Con la hipótesis A y si f es convexa en $\mathbb{R}^n$, la sucesión $\{x^k\}$ converge a una solución óptima del problema (2.1).*

Corolário 2.2 *Con la hipótesis A and si f es diferenciable y pseudoconvexa en $\mathbb{R}^n$, entonces la sucesión $\{x^k\}$ converge a una solución óptima del problema (2.1).*

Cuando los parámetros λ_k satisfacen la condición:

$$\lim_{k \rightarrow +\infty} \lambda_k = 0, \quad \text{con } \lambda_k > 0, \quad (2.14)$$

es posible probar la convergencia global de $\{x^k\}$ a un punto de mínimo, como veremos a seguir.

Teorema 2.5 *Con las hipótesis A' and B', la sucesión $\{x^k\}$ generada por el método del punto proximal, con λ_k satisfaciendo (2.14), converge a una solución optima del problema (2.1).*

Demostración. Sea $x^* \in X^*$ y sea $\bar{x} \in \mathbb{R}^n$ el punto de convergencia de la sucesión $\{x^k\}$, el cual es garantizada por el Teorema 2.4, entonces

$$\lim_{k \rightarrow +\infty} x^k = \bar{x}.$$

Como x^k es un punto solución de (2.3) tenemos

$$f(x^k) + (\lambda_k/2) \|x^k - x^{k-1}\|^2 \leq f(x^*) + (\lambda_k/2) \|x^* - x^{k-1}\|^2.$$

Reescribiendo esta desigualdad

$$(\lambda_k/2) (\|x^k - x^{k-1}\|^2 - \|x^* - x^{k-1}\|^2) \leq f(x^*) - f(x^k).$$

Tomando $k \rightarrow +\infty$ y usando la Proposición 2.2, ítem **b**, obtenemos

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

Ahora daremos una estimación de la distancia del punto inicial a un punto de convergencia de la sucesión $\{x^k\}$.

Sea $P_U : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tal que

$$P_U(x^0) = \arg \min\{\|x - x^0\| : x \in U\}.$$

Note que $P_U(x^0)$ existe ya que U es un conjunto no vacío cerrado, ver Ejemplo 3.8 de Papa Quiroz [17].

Denotemos por

$$\rho(x^0, U) := \inf\{\|x - x^0\| : x \in U\}.$$

Proposición 2.3 *Sea $\{x^k\}$ la sucesión generada por el método de punto proximal y $\bar{x}$ el punto de convergencia, entonces*

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

Prueba. Fijando $x = P_U(x^0)$ en (2.13) obtenemos

$$\|P_U(x^0) - x^k\| \leq \|P_U(x^0) - x^0\|.$$

Tomando $k \rightarrow \infty$

$$\|P_U(x^0) - \bar{x}\| \leq \|P_U(x^0) - x^0\|. \quad (2.15)$$

Ahora, por la desigualdad triangular

$$\begin{aligned} \|x^0 - \bar{x}\|^2 &\leq (\|x^0 - P_U(x^0)\| + \|P_U(x^0) - \bar{x}\|)^2 \\ &\leq 2(\|x^0 - P_U(x^0)\|^2 + \|P_U(x^0) - \bar{x}\|^2) \\ &\leq 2(\|x^0 - P_U(x^0)\|^2 + \|x^0 - P_U(x^0)\|^2) \\ &= 4\|x^0 - P_U(x^0)\|^2 \end{aligned}$$

donde la tercera desigualdad es debido a (2.15). Así, la demostración es completada. ■

Observación 2.2 *Si $\bar{x}$ es un mínimo global de (2.1), por ejemplo cuando f es convexa o diferenciable y pseudoconvexa, el resultado anterior sigue siendo verdadero si sustituimos U por X^* .*

2.5 Caso Sub-Analítico

Denotaremos por $W(x^0)$ y $\text{Crit}f$ el conjunto de puntos de acumulación y puntos críticos de f respectivamente. En esta sección usaremos que los parametros λ_k satisfaciendo:

$$\lambda' < \lambda_k < \bar{\lambda}, \quad (2.16)$$

donde $\lambda' > 0$.

Proposición 2.4 *Sea $\{x^k\}$ la sucesión generada por el método del punto proximal satisfaciendo (2.16). Si las hipótesis A y B son satisfechas, entonces los siguientes enunciados son verdaderos*

a. $\sum_{k=1}^{\infty} \|x^k - x^{k-1}\|^2 < +\infty$

b. $W(x^0) \subset \text{Crit}f$;

Si adicionalmente, $\{x^k\}$ es acotado entonces:

c. f es finito y constante en $W(x^0)$.

d. $W(x^0)$ es connexo, compacto y

$$d(x^k, W(x^0)) \rightarrow 0, k \rightarrow +\infty.$$

Demostración. Inmediato ■.

Como fue observado por Absil et al. [1], la convergencia de toda la sucesión $\{x^k\}$ puede fallar, por eso introducimos la siguiente propiedad.

Propiedad de Lojasiewicz

Para cada $\hat{x} \in \text{Crit}f$, existen ϵ , C y $\theta \in [0, 1)$ tal que

$$|f(x) - f(\hat{x})|^\theta \leq C \|x^*\|,$$

para todo $x \in B(\hat{x}, \epsilon)$ y todo $x^* \in \partial f(x)$.

Hipótesis C. $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$ satisface la propiedad de Lojasiewicz .

Ahora, introducimos un lemma que será usado en la demostración de la convergencia de $\{x^k\}$.

Lemma 2.3 *Asumese que f satisface la propiedad de Lojasiewicz, entonces*

- i. *Si K es un conjunto no vacío conexo de $\text{Crit}f$, entonces f es constante en K .*
- ii. *Ademas, si K es compacto, entonces existen ϵ , C y $\theta \in [0, 1)$ tal que para todo $x \in M$ con $d(x, K) \leq \epsilon$,*

$$|f(x) - f(\hat{x})|^\theta \leq C \|x^*\| ,$$

para todo $x^* \in \partial f(x)$.

Prueba. Ver Attouch and Bolte [2]. ■

Teorema 2.6 *Supongamos que las hipótesis A, B y C son satisfechas. Si $\{x^k\}$ es acotado, entonces*

$$\sum_{k=1}^{\infty} d(x^k, x^{k-1}) < +\infty.$$

En particular, $\{x^k\}$ converge a algún punto de $\text{Crit}f$.

Demostración. Como $\{f(x^k)\}$ es una sucesión limitada inferiormente y no creciente entonces, esta sucesión converge y además

$$f(x^k) \rightarrow \inf f(x^k) \text{ cuando } k \rightarrow +\infty.$$

Defina

$$g(x^k) = f(x^k) - \inf f(x^k)$$

Entonces obtenemos que $g(x^k) \geq 0$ y

$$g(x^k) \rightarrow 0 \text{ cuando } k \rightarrow +\infty.$$

Analizamos dos casos:

Si $x^l = x^{l-1}$ para algún $l > 0$ entonces, de (2.3) tenemos que $0 \in \partial f(x^l)$ y el algoritmo

finaliza en un número finito de iteraciones. Por otro lado, considere $x^k \neq x^{k-1}$ para todo k , entonces $d(x^k, x^{k+1}) > 0$ y por eso $\{g(x^k)\}$ es una sucesión positiva estrictamente decreciente.

Ahora, definimos la función $h : \mathbb{R}^{++} \rightarrow \mathbb{R}$ tal que $h(s) = -s^{1-\theta}$, $\theta \in [0, 1)$. Esta función es diferenciable y estrictamente convexa, por eso

$$h(s) - h(s_0) \geq h'(s_0)(s - s_0) \quad (2.17)$$

Tomando $s_0 = g(x^k)$ y $s = g(x^{k+1})$ en (2.17) tenemos

$$g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta} \geq (1-\theta)g(x^k)^{-\theta}(g(x^k) - g(x^{k+1})).$$

Como $(\lambda_k/2)d^2(x^{k+1}, x^k) \leq g(x^k) - g(x^{k+1})$ entonces

$$g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta} \geq (\lambda_k/2)(1-\theta)g(x^k)^{-\theta}d^2(x^{k+1}, x^k) \quad (2.18)$$

Usando el Lema 2.3 para $K = W(x^0)$ obtenemos que existe $\hat{x} \in W(x^0)$, N_0 , ϵ , C y $\theta \in [0, 1)$ tal que para todo $k \geq N_0$,

$$|f(x^k) - f(\hat{x})|^\theta \leq C\|g^k\|,$$

para cualquier $x^* \in \partial f(x)$. Como f es constante en K , podemos sustituir $f(\hat{x})$ por $f(\bar{x})$, donde $\bar{x}$ es un punto de acumulación de $\{x^k\}$. Así $\forall k \geq N_0$

$$|f(x^k) - f(\bar{x})|^\theta \leq C\|g^k\|.$$

Como $\{f(x^k)\}$ converge a $\inf f(x^k)$ y por lo tanto $\forall k \geq N_0$:

$$0 < g(x^k)^\theta \leq C\lambda_k d(x^k, x^{k-1}) \quad (2.19)$$

Observamos que $\theta \neq 0$. En efecto, si $\theta = 0$ entonces $\forall k \geq N_0 : 1 \leq C\lambda_k d(x^k, x^{k-1})$. Tomando la subsucesión $\{x^{k_j}\}$ de $\{x^k\}$ tal que $x^{k_j} \rightarrow \bar{x}$ obtenemos que $1 \leq 0$. Por lo tanto, $\theta \in (0, 1)$. De (2.19) y como $\theta \neq 0$ obtenemos

$$d(x^k, x^{k-1})^{-1}/(C\lambda_k) \leq g(x^k)^{-\theta}$$

$\forall k \geq N_0$.

Usando la desigualdad anterior en (2.18) obtenemos que existe $M > 0$ tal que $\forall k \geq N_0$:

$$\frac{d(x^{k+1}, x^k)^2}{d(x^k, x^{k-1})} \leq M(g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta}) \quad (2.20)$$

Si $r \in (0, 1)$ y tomando $k \geq N_0$. Si $d(x^{k+1}, x^k) \geq rd(x^k, x^{k-1})$, (2.20) implica que

$$d(x^{k+1}, x^k) \leq (M/r)(g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta}),$$

y por lo tanto $\forall k \geq N_0$:

$$d(x^{k+1}, x^k) \leq rd(x^k, x^{k-1}) + (M/r)(g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta}).$$

Ahora, si $N \geq N_0$ entonces

$$\sum_{k=N_0}^N d(x^{k+1}, x^k) \leq \frac{r}{1-r} d(x^{N_0}, x^{N_0-1}) + (M/(r(1-r)))(g(x^{N_0})^{1-\theta} - g(x^{N+1})^{1-\theta}),$$

y la conclusión sigue del hecho que f es limitado inferiormente. ■

Bibliografia

- [1] P.A. Absil, R. Mahony and B. Andrews, Convergence of the Iterates of Descent Methods for Analytic Cost Function, SIAM J. Optim., 16, 531-547, 2005.
- [2] H. Attouch, and J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Mathematical Programming, 116, 1-2, Ser. B: 5-16, 2009.
- [3] H. Attouch, and M. Teboulle, Regularized Lotka-Volterra dynamical system as continuous proximal-like method in optimization, Journal Optimization Theory and Applications, 121 (2004) 541-580.
- [4] H. Attouch, and A. Soubeyran, Inertia and reactivity in decision making as cognitive variational inequalities, to appear in Journal of Convex Analysis.
- [5] H., Attouch and A. Soubeyran, From procedural rationality to routines: a worthwhile to move approach of satisficing with not too much sacrificing, submitted paper, 2005.
- [6] N. Barron and W. Liu, Calculus of variation l^∞ , Applied Math. Optim. 35 (1997) 237-263.
- [7] Bazaraa M., Sherali H. and Shetty, Non linear programming: theory and algorithms, Jhon Wiley & Sons, Inc., United States of America, 1993.
- [8] R.S. Burachik, L.M. Graña Drummond, A.N. Iusem and B.F. Svaiter, Full convergence of the steepest descent method with inexact line searches, Optimization, 32 (1995) 137-146.

- [9] X. Goudou and J. Munier, The heavy ball with friction method: the quasiconvex case, Submmited, 2005.
- [10] O. Guler, On the convergence of the proximal point algorithm for convex minimization, SIAM J. Crontro and Optimization, 29(2) (1991) 403-419.
- [11] A.N. Iusem and B.F. Svaiter, A proximal regularization of the steepest descent method, RAIRO Recherche Opérationelle/Operation Reseach, 29 (2) (1995) 123-130.
- [12] Iusem, A.N., Métodos de ponto proximal em otimização, 20 Coloquio Brasileiro de Matemática, IMPA, 1995.
- [13] Jein-Shan Chen, A proximal-like algorithm for a class of nonconvex programming. Pacific Journal of Optimization,4: 319-333, 2008.
- [14] J. Gromicho, Quasiconvex optimization and location theory, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1998.
- [15] K. Kiwiel and K. Murty, Convergence of the steepest descent method for minimization quasiconvex functions, JOTA Journal of Optimization Theory and Applications, 89 (1) 1996, 221-226.
- [16] B. Martinet, Régularization, d'inéquations variationelles par approximations successives, Revue Francaise q'Informatique et de Recherche Operationelle, pp. 154-159, 1970.
- [17] E.A. Papa Quiroz, Optimización continua: teoría y ejercicios, Editorial Académica Española, 2017.
- [18] E.A. Papa Quiroz and P. Roberto Oliveira, Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds, Journal of Convex Analysis, 2009.
- [19] E.A. Papa Quiroz E.M. Quispe, and P. Roberto Oliveira, Steepest descent method for quasiconvex minimization on Riemannian manifolds, Submmited, 2006.

- [20] Papa Quiroz, E. A. and Oliveira, P. R. (2006), Classical and logarithmic-quadratic proximal point method for quasiconvex minimization. Optimization Online, Preprint.
- [21] A. Takayama, Mathematical Economics, *2nd* Edition, Cambridge Univ. Press, 1995.
- [22] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control and Optimization, 14 (5) 1976, 877-898.
- [23] R.T. Rockafellar and R. Wets, Variational Analysis. Grundlehren der Mathematischen Wissenschaften, 317, Springer, 1998.
- [24] Souza SS, Oliveira PR, da Cruz Neto JX, Soubeyran A (2010) A proximal method with separable Bregman distance for quasiconvex minimization on the nonnegative orthant. European Journal of Operational Research, 201: 365-376.