

## UNA EXTENSIÓN DEL MÉTODO SUBGRADIENTE PARA FUNCIONES CUASICONVEXAS

*Frank Navarro*<sup>1</sup>, *Tomas Nuñez*<sup>2</sup>

**Resumen:** En el presente trabajo, consideramos el problema de minimizar una función continua, cuasiconvexa y Hölder sobre el conjunto optimal, no necesariamente diferenciable. Para esto utilizamos las direcciones normalizadas del cono normal de los conjuntos de nivel de la función y elegimos los pasos basándonos en el conocimiento del valor óptimo de la función objetivo, también presentamos un ejemplo y su implementación computacional en Matlab.

**Palabras clave:** análisis convexo, optimización no diferenciable, funciones cuasiconvexas, método subgradiente.

**Abstrac:** In this work, we consider the problem of minimizing a quasiconvex, continue and Hölder function on the set optimal, not necessarily differentiable. We use the normalized direction of the normal cone of the set level of function and employ the stepsize rule based in knowledge of the optimal value of the objective function; we also present an example and us computational implementations in Matlab.

**Key words:** analysis convex, nonsmooth optimization, function quasiconvex, subgradient method.

### 1. Introducción

La optimización puede ser dividida en diferenciable y no diferenciable. En el caso de optimización diferenciable se trabaja con problemas donde las funciones poseen derivadas en todos los puntos de su dominio, en cuanto que en optimización no diferenciable se trabaja con problemas cuyas funciones no poseen derivadas en algunos puntos de su dominio. En general, los métodos de optimización son métodos iterativos, los que a partir de un punto inicial, construyen una sucesión de puntos que se aproximen a la solución óptima del problema. En aplicaciones prácticas de optimización como en economía, teoría de control y relajación lagrangeana a menudo se presentan situaciones donde la función objetivo a ser minimizada o maximizada no es necesariamente diferenciable.

Para resolver esta clase de problemas son necesarias técnicas que reemplacen las del cálculo diferencial clásico. En optimización no diferenciable, existen puntos del dominio donde la función objetivo no posee gradiente y en consecuencia tampoco matriz Hessiana. Así, aplicar métodos clásicos para resolver problemas dife-

renciables en problemas no diferenciables se torna inviable. En estos casos, se suele en la mayoría de veces, adaptar los métodos clásicos para funciones no diferenciables.

Por ejemplo, los Métodos Gradientes han sido modificados para funciones no diferenciables, reemplazando la dirección opuesta al gradiente, por la dirección opuesta a un subgradiente, como una dirección de búsqueda para el caso convexo. Los métodos de optimización generalmente utilizan una recurrencia de la forma  $x_{k+1} = x_k + l_k d_k$  donde  $l_k$  es el tamaño de paso y  $d_k$  es una dirección de búsqueda.

Dentro de la clase de métodos para optimización no diferenciable destaca el método del subgradiente; estos métodos, fueron elaborados por Shor en la Unión Soviética en la década de los 60, fueron los primeros métodos para optimización no diferenciable, diseñados para minimizar una función convexa no necesariamente diferenciable, el método básico del subgradiente poseen una estructura simple, de

<sup>1</sup>UNMSM, Facultad de Ciencias Matemáticas, e-mail: nilofrank@hotmail.com

<sup>2</sup>UNMSM, Facultad de Ciencias Matemáticas, e-mail: tnunez@gmail.com

la forma  $x_{k+1} = x_k - l_k d_k$  donde  $d_k$  es un subgradiente de  $f$  en  $x_k$  i.e  $d_k \in \partial f(x_k)$ .

Los métodos subgradientes usan un subgradiente como una dirección de búsqueda que no es necesariamente es una dirección de descenso, así no se puede tener una garantía de la convergencia del método. Para poder obtener resultados de convergencia se elijen los tamaños de paso  $l_k$  de manera conveniente.

Una de estas elecciones es debido a Polyac quien utiliza el valor óptimo del problema  $f^*$ , asumido conocido y elige los pasos como  $l_k = \gamma \frac{f(x_k) - f^*}{\|d_k\|^2}$  para  $0 < \gamma < 2$ , con esta elección prueba la convergencia del método.

Un problema es extender el método del subgradiente a una clase más amplia de funciones que las funciones convexas una extensión natural es al de las funciones cuasiconvexas en este trabajo nos planteamos dicho problema utilizando los pasos dados por Polyac y tomando las direcciones de búsqueda en el cono normal de los conjuntos de nivel de la función, bajo la hipótesis adicional de que la función es Hölder sobre el conjunto optimal.

## 2. Preliminares

**Definición 2.1.** Una sucesión  $(x_k)$  en  $\mathbb{R}^n$  es Fejér convergente en un conjunto  $U \subseteq \mathbb{R}^n$ , si

$$\|x_{k+1} - x\| \leq \|x_k - x\| \quad \text{para cualesquiera } k \in \mathbb{Z}_0^+, x \in U$$

Como consecuencia de la definición anterior tenemos la siguiente proposición.

**Proposición 2.1.** Sea  $(x_k)$  una sucesión Fejér convergente en un conjunto  $U \subseteq \mathbb{R}^n$  entonces  $(x_k)$  es acotada. Si  $x \in U$  es un punto de acumulación de  $(x_k)$  entonces  $\lim_{k \rightarrow \infty} x_k = x$ .

### Prueba

Se sigue de la definición que  $\|x_{k+1} - x\| \leq \|x_0 - x\|$  para toda  $x \in U$  y para toda  $k \in \mathbb{N}$ , entonces  $(x_k)$  es acotada. Ahora sea  $(x_{k_j})$  una subsucesión de  $(x_k)$  tal que  $\lim_{j \rightarrow \infty} x_{k_j} = x$ , tenemos que  $(\|x_k - x\|)$  es una sucesión decreciente de numeros no negativos y posee una subsucesión  $(\|x_{k_j} - x\|)$  que converge a 0 luego  $\lim_{k \rightarrow \infty} \|x_k - x\| = 0$  entonces  $\lim_{k \rightarrow \infty} x_k = x$ .

**Definición 2.2.** Una función  $f : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  se dice que satisface la condición de Hölder con constante  $L > 0$  y grado  $\beta$  en  $x \in C$  si :

$$|f(x) - f(y)| \leq L \|x - y\|^\beta \quad \text{para cualquier } y \in C.$$

En particular si  $\beta = 1$  y se satisface la desigualdad anterior para toda  $x \in C$  entonces  $f$  es lipchitziana en  $C$ .

**Definición 2.3.** Una función  $f : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , con  $C$  convexo, es llamada cuasiconvexa si para todo  $x, y \in C$  se verifica :

$$f(lx + (1-l)y) \leq \max\{f(x), f(y)\} \quad \text{para cualquier } l \in [0, 1].$$

**Teorema 2.1.** Sea  $f : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , donde  $C$  es un conjunto convexo y no vacío. La función  $f$  es cuasiconvexa si y solamente si los conjuntos de nivel  $L_k(f) = \{x \in C : f(x) \leq k\}$  son convexos para todo  $k \in \mathbb{R}$

### Prueba

Supongamos que  $f$  es cuasiconvexa y sean  $x, y \in L_k(f)$  entonces  $\max\{f(x), f(y)\} \leq k$ , sea  $z = lx + (1-l)y \in C$  con  $l \in [0, 1]$ , como  $f$  es cuasiconvexa tenemos  $f(z) \leq \max\{f(x), f(y)\} \leq k$  entonces  $z \in L_k(f)$  así  $L_k(f)$  es convexo.

Ahora supongamos que  $L_k(f)$  es convexo para cada  $k \in \mathbb{R}$ . Sean  $x, y \in C$  entonces  $x, y \in L_{\bar{k}}(f)$  eligiendo  $\bar{k} = \max\{f(x), f(y)\}$ . Sea  $z = lx + (1-l)y \in C$  con  $l \in [0, 1]$  por Hipótesis  $L_{\bar{k}}(f)$  es convexo entonces  $z \in L_{\bar{k}}(f)$  entonces  $f(z) \leq \bar{k} = \max\{f(x), f(y)\}$  así  $f$  es cuasiconvexa.

**Definición 2.4.** Sean  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x_0 \in \mathbb{R}^n$ . Un vector  $s \in \mathbb{R}^n$  tal que:

$$f(x) \geq f(x_0) + \langle s, x - x_0 \rangle, \quad \forall x \in \mathbb{R}^n$$

es llamado un subgradiente de  $f$  en  $x_0$ .

El conjunto de subgradientes de una función  $f$  en el punto  $x_0$  es llamado el subdiferencial de  $f$  en  $x_0$  y se denota por  $\partial f(x_0)$ .

**Definición 2.5.** Sea  $C \subseteq \mathbb{R}^n$  un conjunto convexo y  $x \in C$ . El cono normal de  $x$  en  $C$  es dado por:

$$N_C(x) = \{y \in \mathbb{R}^n; \langle y, z - x \rangle \leq 0, \quad \forall z \in C\}.$$

**Teorema 2.2.** (Teorema de separación) Sean  $C_1, C_2 \subseteq \mathbb{R}^n$  conjuntos convexos y no vacíos tales que  $C_1 \cap C_2 = \emptyset$ . Entonces existen  $a \in \mathbb{R}^n - \{0\}$  y  $c \in \mathbb{R}$  tales que:

$$\langle a, x \rangle \leq c \leq \langle a, y \rangle, \quad \text{para cualesquiera } x \in C_1, y \in C_2.$$

**Prueba**

Ver [1]

### 3. Hipótesis y planteamiento del problema

Consideremos el problema :

$$PI : \min_{x \in C} f(x)$$

Admitiremos que el problema tiene solución, con conjunto optimal  $X^*$  y denotemos con  $f^*$  al valor óptimo;  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es una función continua, cuasiconvexa y satisface la condición de Hölder sobre  $X^*$ , entonces es fácil comprobar que el conjunto optimal  $X^*$  es no vacío, convexo y cerrado.

Definimos el  $\epsilon$ -conjunto de nivel estricto de la función  $f$  dada por

$$G_\epsilon(x) = \{z \in \mathbb{R}^n; f(z) < f(x) - \epsilon\}$$

Por simplicidad denotaremos  $G_0(x) = G(x)$ , notese que  $G_\epsilon(x) \subseteq G(x)$  para todo  $\epsilon \geq 0$ .

Consideremos el conjunto

$$Q_\epsilon(x) = S(0, 1) \cap N_{G_\epsilon(x)}(x)$$

si  $x \notin X^*$  y  $f^* - f(x) + \epsilon < 0$  entonces  $Q_\epsilon(x)$  es no vacío; en efecto, notese primero que si  $b \in N_{G_\epsilon(x)}(x)$  entonces  $\frac{b}{\|b\|} \in N_{G_\epsilon(x)}(x)$ , ahora puesto que,  $x \notin G_\epsilon(x)$  y  $G_\epsilon(x)$  es convexo entonces por el Teorema 2.1 existe  $a \neq 0$  tal que  $\langle a, y \rangle \leq \langle a, x \rangle$ , así  $\langle a, y - x \rangle \leq 0$  para toda  $y \in G_\epsilon(x)$  entonces  $a \in N_{G_\epsilon(x)}(x)$ , luego  $\frac{a}{\|a\|} \in Q_\epsilon(x)$ .

Si  $f$  es una función convexa y  $\epsilon = 0$  se tiene que  $\partial f(x) \subseteq N_{G(x)}(x)$  esta observación nos permite en analogía al método del subgradiente considerar el método iterativo dado por:

$$x_{k+1} = x_k - l_k q_k, \quad q_k \in Q(x_k), \quad l_k > 0$$

### 4. Teoremas de convergencia

El siguiente teorema puede verse como una extensión de la desigualdad que define un subgradiente  $\langle s, x^* - x \rangle + f(x) \leq f^*$  para  $L = 1, \beta = 1, \epsilon = 0$ .

**Teorema 4.1.** Suponga que la función  $f$  satisface la condición de Hölder con grado  $\beta > 0$  y constante  $L$  en  $x^* \in X^*$  sobre el conjunto  $\overline{G}(x)$  para todo  $x \in \mathbb{R}^n \setminus X^*$ , entonces para todo  $\epsilon \geq 0$  tal que  $f(x) - f^* > \epsilon$  tenemos:

$$f(x) - L \langle q, x - x^* \rangle^\beta - \epsilon \leq f^*, \quad \text{para toda } q \in Q_\epsilon(x) \quad (31)$$

### Prueba

Como  $f$  es continua y cuasiconvexa, se tiene que  $G_\epsilon(x)$  es convexo y abierto; veamos que es convexo, sean  $a, b \in G_\epsilon(x)$ , tomemos  $c = la + (1-l)b$  con  $l \in [0, 1]$ , entonces

$$f(c) = f(la + (1-l)b) \leq \max\{f(a), f(b)\} < f(x) - \epsilon$$

esto implica que  $c \in G_\epsilon(x)$  luego  $G_\epsilon(x)$  es convexo. Que es abierto se deduce de inmediato de la continuidad.

Sea  $r = \inf\{\|z - x^*\|; z \in Fr(G_\epsilon(x))\}$  puesto que  $Fr(G_\epsilon(x))$  es cerrado y  $x^* \notin Fr(G_\epsilon(x))$  entonces  $0 < r < \infty^+$ ; por la definición de ínfimo existen una sucesión  $(z_k) \subseteq Fr(G_\epsilon(x))$  y  $(l_k) \subseteq \mathbb{R}^+$  con  $l_k \rightarrow 0$  tal que  $\|z_k - x^*\| \leq r + l_k \quad \forall k \in \mathbb{Z}_0^+$  y puesto que  $z_k \notin G_\epsilon(x)$  se tiene:

$$f(x) - f^* - \epsilon \leq f(z_k) - f^* \leq L \|z_k - x^*\|^\beta \leq L(r + l_k)^\beta$$

Haciendo  $k \rightarrow \infty$ , se tiene:

$$f(x) - f^* \leq Lr^\beta + \epsilon \quad (32)$$

Ahora como  $G_\epsilon(x) \subseteq G(x)$  y por la elección de  $\epsilon \geq 0$  se tiene que  $x^* \in G_\epsilon(x)$ , así  $x^* + zr \in G_\epsilon(x)$  para toda  $z \in S(0, 1)$  tenemos  $\langle q, x^* + zr - x \rangle \leq 0$ , para toda  $q \in Q_\epsilon(x)$ , tomando  $z = q$  se tiene:

$$r \langle q, q \rangle + \langle q, x^* - x \rangle \leq 0$$

así

$$r \leq \langle q, x - x^* \rangle \quad \forall q \in Q_\epsilon(x)$$

Luego, de (2), tenemos  $f(x) - f^* \leq L \langle q, x - x^* \rangle^\beta + \epsilon$ , para cualquier  $q \in Q_\epsilon(x)$ .

### Algoritmo

Consideremos el siguiente algoritmo para el problema  $PI$ , asumiendo conocido el valor óptimo  $f^*$  y tomemos  $\epsilon = 0$ .

Dado un punto inicial  $x_0 \in \mathbb{R}^n \setminus X^*$  construimos una sucesión  $(x_k)$  de la siguiente manera:

$$x_{k+1} = x_k - l_k q_k, \quad q_k \in Q(x_k), \quad l_k > 0$$

donde  $l_k$  es elegido como  $l_k = \gamma \left[ \frac{f(x_k) - f^*}{L} \right]^{\frac{1}{\beta}}$  con  $\gamma \in (0, 2)$ .

**Teorema 4.2.** Suponga que la función  $f$  satisface la condición de Hölder con constante  $L > 0$  y grado  $\beta > 0$  en  $x^* \in X^*$  entonces:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \gamma(2 - \gamma) \left( \frac{f(x_k) - f^*}{L} \right)^{\frac{2}{\beta}} \quad (33)$$

### Prueba

Por la definición del algoritmo tenemos

$$\|x_{k+1} - x^*\|^2 = \|x_k - l_k q_k - x^*\|^2 = \|x_k - x^*\|^2 - 2l_k \langle q_k, x_k - x^* \rangle + l_k^2$$

Usando la definición de  $l_k$  y de (1) tenemos:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\gamma \left[ \frac{f(x_k) - f^*}{L} \right]^{\frac{2}{\beta}} + \gamma^2 \left[ \frac{f(x_k) - f^*}{L} \right]^{\frac{2}{\beta}} =$$

$$\|x_k - x^*\|^2 - \gamma(2 - \gamma) \left[ \frac{f(x_k) - f^*}{L} \right]^{\frac{2}{\beta}}$$

lo que prueba el teorema.

**Corolario 4.1.** Si todas las hipótesis del teorema anterior son satisfechas entonces

i)  $\lim_{k \rightarrow \infty} f(x_k) = f^*$

ii) Los puntos de acumulación de la sucesión  $(x_k)$  pertenecen a  $X^*$  y

$\lim_{k \rightarrow \infty} x_k = z \in X^*$

### Prueba

Se sigue de (3), que la sucesión  $(x_k)$  satisface

$$\|x_{k+1} - x^*\| \leq \|x_k - x^*\|$$

así  $(x_k)$  es Fejér convergente en  $X^*$ , luego es acotada entonces tiene un punto de acumulación. Además de (3) tenemos :

$$\gamma(2 - \gamma) \sum_{i=0}^n \left[ \frac{f(x_i) - f^*}{L} \right]^{\frac{2}{\beta}} \leq \|x_0 - x^*\|^2$$

entonces tenemos

$$\sum_{i=0}^{\infty} (f(x_i) - f^*)^{\frac{2}{\beta}} < \infty^+$$

luego

$$\lim_{k \rightarrow \infty} f(x_k) = f^*$$

Ahora, si  $z$  es un punto de acumulación de  $(x_k)$  entonces por continuidad  $f(z) = f^*$  entonces  $z \in X^*$ , así por el teorema se tiene

$$\lim_{k \rightarrow \infty} x_k = z$$

## 5. Implementación numérica

Consideremos la función:

$$g(x) = \begin{cases} -2x - 1 & x \leq -1 \\ -x^3 & -1 \leq x \leq 1 \\ \sqrt{x} - 2 & 1 \leq x \leq 9 \\ x - 8 & 9 \leq x \end{cases}$$

La función anterior es cuasiconvexa pues todos sus conjuntos de nivel son convexos. Se tiene que  $X^* = \{1\}$  y  $f^* = 1$ , la función satisface la condición de Hölder sobre  $X^*$  con  $L = 3$  y  $\beta = 1$ , implementamos el algoritmo :

$$x_{k+1} = x_k - l_k q_k$$

con  $l_k = \gamma \left[ \frac{f(x_k) - f^*}{L} \right]^{\frac{1}{\beta}}$  donde  $\gamma \in (0, 2)$  y  $q_k \in Q(x_k)$

Nótese que esta función no es diferenciable en  $x = -1$ ,  $x = 1$  y  $x = 9$ , también notamos que si  $x < 1$  entonces  $Q(x) = \{-1\}$  y si  $x > 1$  entonces  $Q(x) = \{1\}$ .

La siguiente tabla muestra la iteración del método para  $\gamma = 0,4$ , tomando como punto inicial  $x_0 = 4$ .

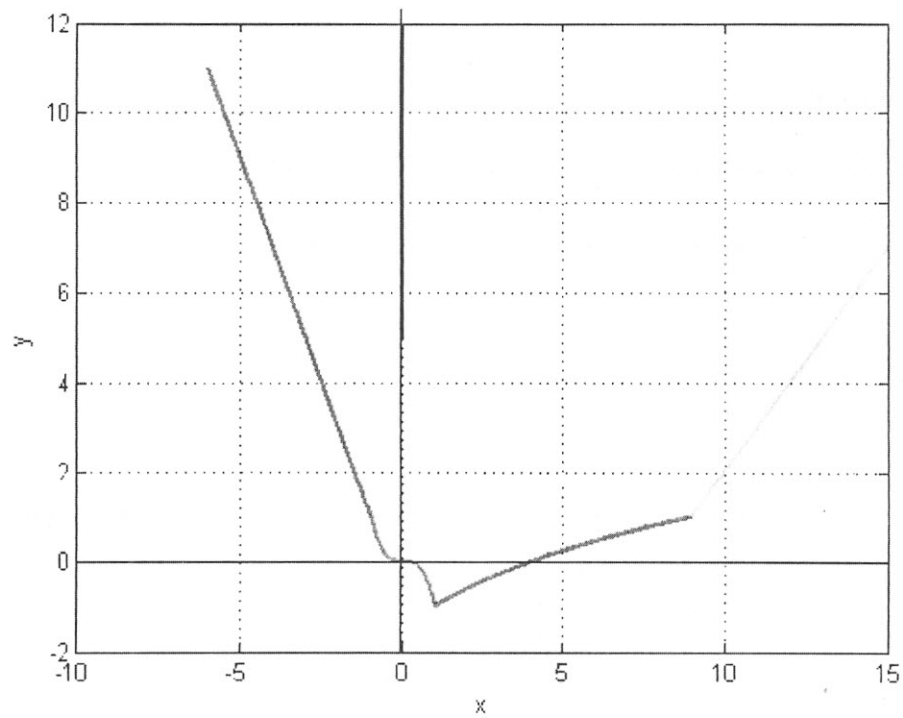


Figura 1: Gráfica de la función cuasiconvexa  $y=g(x)$

iteración $k$	$x_k$	$f(x_k)$
0	4	0
1	3.8667	-0.0336
2	3.7378	-0.0667
8	3.1581	-0.2229
15	2.5121	-0.4150
20	2.1543	-0.5323
30	1.6482	-0.7162
40	1.3493	-0.8384
50	1.1827	-0.9125
60	1.0938	-0.9542
70	1.0121	-0.9940

La siguiente tabla muestra la iteración del método para  $\gamma = 1$ , tomando como punto inicial  $x_0 = 3$ .

iteración $k$	$x_k$	$f(x_k)$
0	3	-0.2679
1	2.7560	-0.3399
2	2.5359	-0.4075
8	1.2002	-0.9045
10	1.1414	-0.9316
15	1.0582	-0.9713
20	1.0236	-0.9882
25	1.0095	-0.9952
30	1.0038	-0.9981
35	1.0015	-0.9992
40	1.0006	-0.9997
54	1	-1

La siguiente tabla muestra la iteración del método para  $\gamma = 1,7$ , tomando como punto inicial  $x_0 = 2,5$ .

iteración $k$	$x_k$	$f(x_k)$
0	2.5	-0.4189
1	2.1707	-0.5267
2	1.9025	-0.6207
8	1.0156	-0.9922
10	1.0080	-0.9960
12	1.0041	-0.9979
16	1.0011	-0.9995
23	1.0001	-0.9999
26	1	-1

## 5. Conclusiones

Se ha establecido la convergencia de un algoritmo para minimizar una función cuasiconvexa no necesariamente diferenciable, utilizando las direcciones normalizadas del cono normal de los conjuntos de nivel de la función y eligiendo los pasos, basados en el conocimiento priori del valor óptimo de la función objetivo. En la parte de implementación computacional existe cierta dificultad de encontrar la constante  $L$ , el grado  $\beta$  y el cono normal de los conjuntos de nivel para una función objetivo en general; por otra parte los experimentos computacionales muestran que para  $1 < \gamma < 2$  se obtiene una mejor aproximación al óptimo en un número menor de iteraciones.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] Dimitri P.& Bertsekas, **Convex Analysis and Optimization**, Athena Scientific, 2003.
- [2] Leonard D. Berkovitz, **Convexity and Optimization in  $\mathbb{R}^n$** , World Scientific, 1992.
- [3] N.Z. Shor, **Minimization Methods for Non-Differentiable Function**, Naukova Dumka, Kiev 1979.
- [4] Marko M Makela & Pekka Neittaanmaki, **Nonsmooth optimization**, World Scientific, 1992.
- [5] Luc D.T, **Characterizations of quasiconvex functions**, Bull.Astral, Match.Soc; 48,393-406, 1993.
- [6] Berhanu Guta, **Subgradient Optimization Methods in Integer Programming with an Application to a Radiation Therapy Problem**, University of Kaiserslautern, Ph.D Thesis, September- 2003.