

Algoritmo abstracto de ϵ -descenso generalizado

*Estéfany Castillo Ventura*¹ y *Erik Alex Papa Quiroz*²

Resumen: Dado el problema de minimizar una función posiblemente no convexa y no suave en el espacio euclidiano, presentamos un algoritmo abstracto de ϵ -descenso generalizado. Este algoritmo está motivado por la convergencia abstracta de métodos de descenso demostrada en Attouch et al. sección 2 (Math Program Ser A, 137: 91–129, 2013) con una diferencia esencial, consideramos errores escalares en cada aproximación. Como resultado obtenemos, que todo punto de acumulación de las sucesiones generadas por algoritmos que cumplen con las condiciones del algoritmo abstracto son puntos críticos límite generalizados.

Palabras clave: Optimización no suave, optimización no convexa, función coerciva, métodos de descenso, error relativo, errores escalares

Abstract generalized epsilon-descent algorithm

Abstract: Given the problem of minimizing a possibly non-convex and non-smooth function in Euclidean space, we present an abstract generalized ϵ -descent algorithm. This algorithm is motivated by the abstract convergence of descent methods introduced by Attouch et al. section 2 (Math Program Ser A, 137: 91-129, 2013) with one essential difference, we consider scalar errors in each approximation. As a result, we obtain that all accumulation points of the sequences generated by algorithms satisfying the conditions of the abstract algorithm are generalized critical limit points.

Keywords: Non-smooth optimization, non-convex optimization, coercive function, descent methods, relative error, scalar errors

Recibido: 21/09/2022. *Aceptado:* 20/10/2022. *Publicado online:* 30/12/2022.

¹UNMSM, Facultad de Ciencias Matemáticas. e-mail: estefany.castillo@unmsm.edu.pe

²UNMSM, Facultad de Ciencias Matemáticas, e-mail: epapaq@unmsm.edu.pe

1. Introducción

La Optimización Matemática esta bastante avanzada en cuanto al estudio de optimizar funciones convexas, y más aún, si estas son suaves; la motivación viene al estudiar funciones que no son ninguna de estas dos cosas, entre otras razones porque eventualmente aparecen funciones, que surgen de modelar aspectos de la vida real, que no son ni convexas ni suaves. (ver para teoría del consumidor [18], [5], [21], [6], dinámica de grupos [11], [7])

De los diversos métodos desarrollados en la Optimización Matemática, elegimos los métodos de descenso, que en esencia son algoritmos que producen sucesiones decrecientes con la esperanza que converjan a mínimos de una función.

Cuando las funciones objetivo no son convexas (o cuasiconvexas) generalmente los métodos de descenso proporcionan sucesiones que exhiben comportamientos altamente oscilatorios. Probablemente, este fenómeno fue observado por primera vez por Curry (ver [10]); en el marco de las ecuaciones diferenciales ocurren comportamientos similares, en [15] se proporciona una curva acotada no convergente de un sistema de gradiente bidimensional de una función C^∞ , este ejemplo fue adaptado en [1] a métodos de gradiente. Para sortear tales comportamientos nos vemos obligados a trabajar con funciones que poseen ciertas estructuras, por ejemplo, la propiedad de Kurdyka-Lojasiewicz, la coercitividad, entre otras.

Nos enfrentamos entonces al problema de minimizar una función propia semicontinua inferior no suave y no convexa $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ mediante métodos de descenso. Este problema fue abordado, para funciones que tienen la propiedad de Kurdyka-Lojasiewicz, en un marco unificado para métodos de descenso en Attouch et. al. (2013) [3] donde los autores presentaron un análisis de convergencia para métodos de descenso que proporcionan sucesiones que satisfacen una condición de descenso suficiente y tolerancia en el error relativo, que luego se especializa para métodos de descenso inexactos ya conocidos.

Ahora bien, cuando implementamos algoritmos en una computadora, se vuelve necesario considerar errores que pueden atribuirse a los datos de entrada, al propio computo o cálculo, y de la misma manera, cuando surge la necesidad de construir funciones que modelen el comportamiento humano se debe considerar los márgenes de error que surgen de la toma de decisiones humanas. De esta manera justificamos nuestra consideración de errores escalares en cada aproximación.

Algunos de los algoritmos de descenso que consideran los errores son por ejemplo:

- Frankel, Garrigos y Peypouquet (2015) [12] se amplió el trabajo realizado en Attouch et. al. (2013) [3] construyendo un método abstracto de descenso que considera errores escalares en la condición del error relativo, demostrando convergencia a puntos críticos de la función objetivo cuando ésta cumple con la propiedad de Kurdyka-Lojasiewicz, incluso obtienen nuevos resultados para tasas de convergencia.

- Yaohua, Xiaoqi, Chee (2015) [13] en el cual consideran un algoritmo genérico de subgradiente inexacto para resolver un problema de optimización con restricciones cuasiconvexas no diferenciable, donde la inexactitud proviene de errores de cálculo y ruido, que provienen de consideraciones y aplicaciones prácticas.

- Attouch y Soubeyran (2011) [5] desarrollaron un modelo para la toma de decisiones de la "vida real" que es un proceso de decisión incremental "Un enfoque para descubrir qué vale la pena mover para satisfacer objetivos sin sacrificar demasiado". En este contexto, las características de optimización del proceso de decisión son modeladas por un nuevo algoritmo denominado algoritmo proximal de búsqueda local que en esencia es la combinación de algoritmos de búsqueda local con los algoritmos proximales e incluyen parámetros de motivación como de sacrificio a modo de errores.

- Sun, Tang y Dongsheng (2022) [22] presentan una comprensión del algoritmo de aprendizaje en computadoras con baja precisión en el calculo de decimales pues debido a que los dominios

donde se trabajan no son continuos como se considera en los estudios teóricos, se producen errores en computadoras de baja precisión. Por ello se investigan los rendimientos de tres descensos de gradiente con el dominio flotante cuando la función objetivo es uniforme.

- Papa Quiroz, Soubeyran y Oliveira (2022) [18] en el cual introducen el algoritmo proximal inexacto con cuasi distancia donde consideran dos tipos de errores, escalares y vectoriales, y demuestran que el algoritmo converge a puntos críticos para funciones semicontinuas inferiores propias y coercivas.

Por lo mencionado anteriormente, en este trabajo, para minimizar una función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ extendida, propia y semicontinua inferior, presentamos un algoritmo abstracto de descenso que considera errores escalares en cada aproximación, tanto en la condición de descenso suficiente como en la condición del error relativo, de la siguiente manera:

Proponemos el algoritmo que genere una sucesión $(x^k)_{k \in \mathbb{N}}$ que cumpla con las siguientes condiciones:

- Para cada $k \in \mathbb{N}$, $f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \leq f(x^k) + \epsilon_{k+1}$
- Para cada $k \in \mathbb{N}$, existe $w^{k+1} \in \partial_{\epsilon_{k+1}} f(x^{k+1})$ tal que $\|w^{k+1}\| \leq b\|x^{k+1} - x^k\| + \epsilon_{k+1}$

dónde a, b son constantes positivas y $\partial_{\epsilon_{k+1}} f(x^{k+1})$ denota el ϵ -subdiferencial límite de Fréchet (Para más detalles de este subdiferencial, consultar [14]). La primera condición, pretende modelar una propiedad de descenso en la cual se involucra una medida y un criterio de error de disminución suficiente. La segunda condición se origina del hecho bien conocido de que la mayoría de los algoritmos en optimización son generados por una sucesión infinita de subproblemas que involucran procesos de minimización exactos o inexactos. Este es el caso de los métodos de gradiente, el método de Newton, el método de división hacia adelante y hacia atrás, el método de Gauss-Seidel, los métodos proximales, como lo veremos en la Sección 4. Esta criterio refleja las condiciones de optimalidad relativamente inexactas para tales subproblemas de minimización controladas por un criterio de error.

El artículo está organizado de la siguiente manera: en la Sección 2, presentamos algunos conceptos necesarios para el desarrollo de la investigación; en la Sección 3, presentamos el algoritmo abstracto de ϵ -descenso generalizado; en la Sección 4, mostramos especializaciones del algoritmo en métodos de descenso ya conocidos (Subsección 4.1 método de gradiente, Subsección 4.2 método proximal, Subsección 4.3. algoritmo división hacia adelante y hacia atrás, Subsección 4.4. algoritmo de regularización de Gauss Seidel y en la Subsección 4.5. presentamos un ejemplo del método de división hacia adelante y hacia atrás); en la Sección 5, demostramos que todo punto de acumulación (si existen) de las sucesiones generadas por algoritmos que cumplen con las condiciones de nuestro algoritmo abstracto son puntos críticos límite generalizados de la función objetivo, además demostramos que cuando la función es coerciva las sucesiones son acotadas. y en la última sección presentamos nuestras conclusiones.

2. Conceptos básicos

En este artículo \mathbb{R}^n denota el espacio Euclideo, esto es, el espacio vectorial real n -dimensional dotado con la norma $\|\cdot\|$ inducida por el producto interno canónico $\langle \cdot, \cdot \rangle$, esto es, $\|x\| := \langle x, x \rangle^{1/2}$

Definición 2.1. Dada una función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$

1. El dominio de f es definido por $dom(f) := \{x \in \mathbb{R}^n : f(x) < +\infty\}$
2. f es propia si:
 - a) $dom(f) \neq \emptyset$;
 - b) $\forall x \in dom(f) : f(x) > -\infty$
3. f es semicontinua inferior en x^* si para toda sucesión $\{x^l\} \subset \mathbb{R}^n$ tal que $\lim_{l \rightarrow +\infty} x^l = x^*$ se obtiene

$$f(\bar{x}) \leq \liminf_{l \rightarrow +\infty} f(x^l)$$

La función f es semicontinua inferior (en \mathbb{R}^n) si es semicontinua inferior en todo punto de \mathbb{R}^n .

2.1. Función coerciva

Definición 2.2. Decimos que una sucesión $\{x^k\}$ en X es crítica (en relación al conjunto X) si: $\lim_{k \rightarrow +\infty} \|x^k\| = +\infty \vee \lim_{k \rightarrow +\infty} x^k = \bar{x} \in \bar{X} \setminus X$, donde la notación $\bar{x} \in \bar{X} \setminus X$ significa que $\bar{x} \in \bar{X}$ y $x \notin X$.

La función $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es llamada coerciva en X si para toda sucesión crítica $\{x^k\}$ se tiene:

$$\limsup_{k \rightarrow +\infty} f(x^k) = +\infty$$

Proposición 2.1. ([16] Corolario 3.1, página 38) Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ una función propia y coerciva, entonces para todo $c \in \mathbb{R}$, el conjunto de nivel $L_f(c) = \{x \in \mathbb{R}^n : f(x) \leq c\}$ es acotado.

2.2. Subdiferencial de Fréchet y subdiferenciales limite

Definición 2.3. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia.

- (a) Para cada $x \in dom(f)$, el subdiferencial de Fréchet de f en x , denotado por $\hat{\partial}f(x)$, es el conjunto de vectores $v \in \mathbb{R}^n$ tal que:

$$f(y) \geq f(x) + \langle v, y - x \rangle + o(\|y - x\|) \text{ donde } \lim_{y \rightarrow x} \frac{o(\|y - x\|)}{\|y - x\|} = 0$$

O equivalentemente, $\hat{\partial}f(x) := \{v \in \mathbb{R}^n : \liminf_{y \neq x, y \rightarrow x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|} \geq 0\}$.

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

- (b) El subdiferencial limite de f en $x \in \mathbb{R}^n$, denotado por $\partial f(x)$, es el conjunto de vectores $v \in \mathbb{R}^n$ tal que:

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

- (c) Para cada $x \in \text{dom}(f)$, el ϵ -subdiferencial de Fréchet de f en x , denotado por $\hat{\partial}_\epsilon f(x)$, es el conjunto de vectores $v \in \mathbb{R}^n$ tal que:

$$f(y) \geq f(x) + \langle v, y - x \rangle - \epsilon + o(\|y - x\|) \text{ donde } \lim_{y \rightarrow x} \frac{o(\|y - x\|)}{\|y - x\|} = 0$$

O equivalentemente, $\hat{\partial}_\epsilon f(x) := \{v \in \mathbb{R}^n : \liminf_{y \neq x, y \rightarrow x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|} \geq -\epsilon\}$.

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

- (d) El ϵ -subdiferencial limite de f en $x \in \mathbb{R}^n$, denotado por $\partial_\epsilon f(x)$, es el conjunto de vectores $v \in \mathbb{R}^n$ tal que:

$$\partial_\epsilon f(x) := \{v \in \mathbb{R}^n : \exists x^l \rightarrow x, f(x^l) \rightarrow f(x), v^l \in \hat{\partial}_\epsilon f(x^l) \text{ y } v^l \rightarrow v\}$$

- (e) El subdiferencial limite generalizado de f en $x \in \mathbb{R}^n$, denotado por $\bar{\partial} f(x)$, es el conjunto de vectores $v \in \mathbb{R}^n$ tal que:

$$\bar{\partial} f(x) := \{v \in \mathbb{R}^n : \exists x^l \rightarrow x, f(x^l) \rightarrow f(x), v^l \in \partial_{\epsilon_l} f(x^l) \text{ y } v^l \rightarrow v\}$$

Observación 2.1. Cuando $\epsilon = 0$ los conjuntos en (c) y (d), denotan a $\hat{\partial} f(x)$ y $\partial f(x)$ respectivamente.

Observación 2.2. Para una función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ y un punto \bar{x} donde $f(\bar{x}) < +\infty$, los conjuntos $\partial f(\bar{x})$ y $\hat{\partial} f(\bar{x})$ son cerrados, $\hat{\partial} f(\bar{x})$ es convexo y $\hat{\partial} f(\bar{x}) \subset \partial f(\bar{x})$.

Proposición 2.2. ([20] Teorema 10.1, página 422) Si una función propia $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ tiene un punto mínimo local $\bar{x} \in \text{dom}(f)$, entonces $0 \in \hat{\partial} f(\bar{x})$.

Proposición 2.3. ([20] página 302) Si una función propia $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ es diferenciable en \bar{x} , entonces $\hat{\partial} f(\bar{x}) = \{\nabla f(\bar{x})\}$.

Proposición 2.4. ([14] página 73) Sean $f, g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ entonces:

1. $\hat{\partial}_\epsilon f(x)$ es vacío si $x \notin \text{dom}(f)$
2. $\hat{\partial}_{\epsilon_1} f(x) + \hat{\partial}_{\epsilon_2} g(x) \subset \hat{\partial}_{\epsilon_1 + \epsilon_2} (f + g)(x)$
3. $\hat{\partial}_{\epsilon_1} f(x) \subset \hat{\partial}_{\epsilon_2} f(x)$, si $\epsilon_1 \leq \epsilon_2$

Proposición 2.5. ([14] Proposición 2.8, página 77) Sea X un espacio de Banach y suponga que $f : X \rightarrow \mathbb{R}$ es continuamente Fréchet diferenciable en x con derivada de Fréchet $Df(x)$. Entonces

$$\partial_\epsilon f(x) = Df(x) + \epsilon B^*$$

donde B^* es la bola en el dual topológico de X .

Proposición 2.6. ([20] Ejercicio 8.8. (c) página 304) Si $f = g + f_0$, con g finito en \bar{x} y f_0 suave en una vecindad de \bar{x} , entonces $\hat{\partial} f(\bar{x}) = \hat{\partial} g(\bar{x}) + \nabla f_0(\bar{x})$ y $\partial f(\bar{x}) = \partial g(\bar{x}) + \nabla f_0(\bar{x})$

2.3. Lema de descenso

Lema 2.7. *Lema de descenso*

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función y C un subconjunto convexo de \mathbb{R}^n con interior no vacío. Asumiendo que f es de clase C^1 sobre una vecindad de cada punto de C y que ∇f es L -lipschitz continua sobre C . Entonces, para cualquier par de puntos $x, u \in C$,

$$f(u) \leq f(x) + \langle \nabla f(x), u - x \rangle + \frac{L}{2} \|u - x\|^2 \quad (1)$$

Demostración. Sabemos que $G(a) - G(b) = \int_a^b g(t) dt$ donde $g = G'$

Hacemos $g(t) = \langle \nabla f(x + t(u - x)), u - x \rangle$

Consideramos la integral definida de 0 a 1, entonces $G(b) = G(1) = f(u)$ y $G(a) = G(0) = f(x)$. Luego:

$$f(u) - f(x) = \int_0^1 \langle \nabla f(x + t(u - x)), u - x \rangle dt$$

$$f(u) - f(x) = \int_0^1 \langle \nabla f(x + t(u - x)) + \nabla f(x) - \nabla f(x), u - x \rangle dt$$

$$f(u) - f(x) = \langle \nabla f(x), u - x \rangle + \int_0^1 \langle \nabla f(x + t(u - x)) - \nabla f(x), u - x \rangle dt$$

$$f(u) - f(x) - \langle \nabla f(x), u - x \rangle = \int_0^1 \langle \nabla f(x + t(u - x)) - \nabla f(x), u - x \rangle dt$$

entonces:

$$f(u) - f(x) - \langle \nabla f(x), u - x \rangle \leq |f(u) - f(x) - \langle \nabla f(x), u - x \rangle| = \left| \int_0^1 \langle \nabla f(x + t(u - x)) - \nabla f(x), u - x \rangle dt \right|$$

$$\leq \int_0^1 |\langle \nabla f(x + t(u - x)) - \nabla f(x), u - x \rangle| dt \leq \int_0^1 \|\nabla f(x + t(u - x)) - \nabla f(x)\| \|u - x\| dt$$

$$\leq L \int_0^1 \|u - x\| dt = \frac{L}{2} \|u - x\|^2$$

$$\implies f(u) \leq f(x) + \langle \nabla f(x), u - x \rangle + \frac{L}{2} \|u - x\|^2$$

□

3. Algoritmo Abstracto de epsilon-descenso

En esta sección presentamos el algoritmo abstracto motivado por el artículo de Attouch et al. (2013) [3], construido a partir de las propiedades de descenso suficiente y tolerancia al error relativo con la diferencia que en este trabajo consideramos los errores escalares cometidos en cada iteración.

Algoritmo 1 Algoritmo abstracto de ϵ -descenso generalizado

Paso 1 Elegir $x^0 \in \mathbb{R}^n$, $a, b \in \mathbb{R}$ constantes positivas y una sucesión no negativa $\{\epsilon_{k+1}\} \subset \mathbb{R}$ tal que:

$$\sum_{k=0}^{\infty} \epsilon_{k+1} < \infty \quad (2)$$

Paso 2 Para $k = 0, 1, 2, \dots$ a partir de x^k calculamos x^{k+1} tal que:

$$f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \leq f(x^k) + \epsilon_{k+1} \quad (3)$$

$$w^{k+1} \in \partial_{\epsilon_{k+1}} f(x^{k+1}) \quad (4)$$

donde

$$\|w^{k+1}\| \leq b\|x^{k+1} - x^k\| + \epsilon_{k+1} \quad (5)$$

Paso 3 Si $x^k = x^{k+1}$ o $0 \in \partial f(x^{k+1})$, parar. Caso contrario, regresar al paso 2.

Este algoritmo extiende el trabajo en Attouch et al. (2013) [3] para el caso $\epsilon_k = 0$ e incluye algunos algoritmos inexactos de descenso, como lo veremos en la siguiente sección.

4. Algoritmos con la propiedad épsilon-descenso

4.1. Método inexacto de gradiente

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función de clase C^1 cuya gradiente es continua y L -lipschitz. Consideramos el siguiente algoritmo.

Algoritmo 2

Paso 1 Elegir constantes positivas a, b con $a > L$. Fijar un punto inicial $x^0 \in \mathbb{R}^n$ y una sucesión de números reales positivos $\{\epsilon_{k+1}\}$ tal que:

$$\sum_{k=0}^{\infty} \epsilon_{k+1} < \infty \quad (6)$$

Paso 2 Para $k = 0, 1, 2, \dots$ dado $x^k \in \mathbb{R}^n$, encontrar $x^{k+1} \in \mathbb{R}^n$ tal que:

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{a}{2} \|x^{k+1} - x^k\|^2 \leq \epsilon_{k+1} \quad (7)$$

$$\|\nabla f(x^k)\| \leq b\|x^{k+1} - x^k\| \quad (8)$$

Paso 3 Si $x^k = x^{k+1}$ o $\nabla f(x^{k+1}) = 0$, parar. Caso contrario, regresar al paso 2.

La condición (2) del algoritmo abstracto se obtiene directamente de (6). Para la condición (3) aplicaremos el Lema 2.7 en los puntos $u = x^{k+1}$ y $x = x^k$

$$f(x^{k+1}) - f(x^k) \leq \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2} \|x^{k+1} - x^k\|^2$$

sumando en ambos miembros $\frac{a}{2}\|x^{k+1} - x^k\|^2$ tenemos:

$$\begin{aligned} f(x^{k+1}) - f(x^k) + \frac{a}{2}\|x^{k+1} - x^k\|^2 &\leq \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 + \frac{a}{2}\|x^{k+1} - x^k\|^2 \\ &\leq \epsilon_{k+1} + \frac{L}{2}\|x^{k+1} - x^k\|^2 \end{aligned}$$

entonces:

$$f(x^{k+1}) - f(x^k) + \left(\frac{a-L}{2}\right)\|x^{k+1} - x^k\|^2 \leq \epsilon_{k+1}$$

Como $a > L$ tenemos (3).

De la Proposición 2.5 tenemos que:

$$\partial_{\epsilon_{k+1}} f(x^{k+1}) = \{\nabla h(x^{k+1}) + \epsilon_{k+1}y \mid y \in B^*\}$$

Luego $w^{k+1} = \nabla h(x^{k+1}) + \epsilon_{k+1}y \in \partial_{\epsilon_{k+1}} f(x^{k+1})$, donde $y \in B^*$ y como ∇f es L -lipschitz y de la desigualdad triangular obtenemos lo siguiente:

$$\begin{aligned} \|w^{k+1}\| &= \|\nabla f(x^{k+1}) + \epsilon_{k+1}y\| \leq \|\nabla f(x^{k+1})\| + \|\epsilon_{k+1}y\| \\ &\leq \|\nabla f(x^{k+1}) - \nabla f(x^k)\| + \|\nabla f(x^k)\| + |\epsilon_{k+1}|\|y\| \\ &\leq (L+b)\|x^{k+1} - x^k\| + |\epsilon_{k+1}| = (L+b)\|x^{k+1} - x^k\| + \epsilon_{k+1} \end{aligned}$$

cumplíendose la condición (5) de nuestro algoritmo abstracto.

4.2. Método proximal inexacto

4.2.1. Un tipo de algoritmo proximal inexacto

El siguiente algoritmo extiende el algoritmo introducido en Attouch et al. (2013) Sección 4 [3] al considerar el error escalar en la condición de descenso.

Algoritmo 3 Algoritmo proximal inexacto

Paso 1 Elegir $x^0 \in \mathbb{R}^n$, $0 < \underline{\lambda} \leq \bar{\lambda} < \infty$, $0 \leq \sigma < 1$, $0 < \theta \leq 1$ y una sucesión de números reales positivos $\{\epsilon_{k+1}\}$ tal que:

$$\sum_{k=0}^{\infty} \epsilon_{k+1} < \infty \tag{9}$$

Paso 2 Para $k = 0, 1, 2, \dots$, elegir $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$, dado $x^k \in \mathbb{R}^n$, encontrar $x^{k+1} \in \mathbb{R}^n$ y $w^{k+1} \in \mathbb{R}^n$ tal que:

$$f(x^{k+1}) + \frac{\theta}{2\lambda_k}\|x^k - x^{k+1}\|^2 \leq f(x^k) + \epsilon_{k+1} \tag{10}$$

$$w^{k+1} \in \partial_{\epsilon_{k+1}} f(x^{k+1}) \tag{11}$$

$$\|\lambda_k w^{k+1} + x^{k+1} - x^k\|^2 \leq \sigma(\|\lambda_k w^{k+1}\|^2 + \|x^{k+1} - x^k\|^2) \tag{12}$$

Paso 3 Si $x^k = x^{k+1}$ o $0 \in \partial f(x^{k+1})$, parar. Caso contrario, regresar al paso 2.

Las condiciones (2), (3) y (4) del algoritmo 1 las obtenemos directamente de (9), (10) y (11). Para conseguir la condición (5) vamos a utilizar el siguiente lema.

Lema 4.1. ([3] Lema 4.1. página 113) Sea $\sigma \in (0, 1]$. Si $x, y \in \mathbb{R}^n$, y

$$\|x + y\|^2 \leq \sigma(\|x\|^2 + \|y\|^2)$$

entonces

$$\left(\frac{1 - \sigma}{2}\right) (\|x\|^2 + \|y\|^2) \leq -\langle x, y \rangle$$

Asumiendo además que $\sigma \in (0, 1)$

$$\left(\frac{1 - \sqrt{1 - (1 - \sigma)^2}}{1 - \sigma}\right) \|y\| \leq \|x\| \leq \left(\frac{1 + \sqrt{1 - (1 - \sigma)^2}}{1 - \sigma}\right) \|y\|$$

Aplicando el Lema 4.1 a la desigualdad (11) haciendo $x = \lambda_k w^{k+1}$ y $y = \|x^{k+1} - x^k\|$ tenemos que, para algún número positivo $b > 0$

$$\|\lambda_k w^{k+1}\| \leq b \|x^{k+1} - x^k\|$$

entonces, como $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$:

$$\|w^{k+1}\| \leq \frac{b}{\lambda_k} \|x^{k+1} - x^k\| \leq \frac{b}{\underline{\lambda}} \|x^{k+1} - x^k\|$$

y queda demostrado que se cumple la condición (5).

4.2.2. Algoritmo proximal inexacto con cuasi distancias

En Papa Quiroz et al. (2022) [18] desarrollan el denominado algoritmo proximal inexacto con cuasi-distancias para funciones semicontinuas inferiores, con el objetivo de resolver el problema de minimizar una función $f : \mathbb{R} \rightarrow \mathbb{R}^n \cup \{\infty\}$ propia semicontinua inferior y coerciva, dada una cuasi-distancia $q : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ tal que exista β que satisfice $q(x, y) \leq \beta \|x - y\|$.

Algoritmo 4 Algoritmo proximal inexacto con cuasi-distancias

Paso 1 Elegir una cuasi-distancia $q(\cdot, \cdot)$ definida sobre \mathbb{R}^n , un punto inicial $x^0 \in \mathbb{R}^n$, un número real $\sigma > 0$ y dos sucesiones de números reales positivos $\{\lambda_{k+1}\}$ y $\{\epsilon_{k+1}\}$ tal que:

$$\sum_{i \in \mathbb{N}} \epsilon_{i+1} < \infty \quad (13)$$

Paso 2 Para $k = 0, 1, 2, \dots$ dado $x^k \in \mathbb{R}^n$, encontrar $x^{k+1} \in \mathbb{R}^n$ y $\varphi^{k+1} \in \mathbb{R}^n$ tal que:

$$f(x^{k+1}) \leq f(x^k) - \frac{\lambda_{k+1}}{2} q^2(x^k, x^{k+1}) + \epsilon_{k+1} \quad (14)$$

$$\varphi^{k+1} \in \partial_{\epsilon_{k+1}} f(x^{k+1}) \quad (15)$$

$$\|e^{k+1}\| \leq \sigma q(x^k, x^{k+1}) \quad (16)$$

donde

$$\begin{aligned} e^{k+1} &= \varphi^{k+1} + \lambda_{k+1} q(x^k, x^{k+1}) \psi^{k+1} \\ \psi^{k+1} &\in \partial q(x^k, \cdot)(x^{k+1}) \end{aligned} \quad (17)$$

Paso 3 Si $x^k = x^{k+1}$ o $0 \in \partial f(x^{k+1})$, parar. Caso contrario, regresar al paso 2.

De la condición (17) y la desigualdad triangular tenemos:

$$\begin{aligned} \|\varphi^{k+1}\| &\leq \|e^{k+1} - \lambda_{k+1} q(x^k, x^{k+1}) \psi^{k+1}\| \leq \|e^{k+1}\| + \|\lambda_{k+1} q(x^k, x^{k+1}) \psi^{k+1}\| \\ &\leq \sigma q(x^k, x^{k+1}) + \lambda_{k+1} q(x^k, x^{k+1}) \|\psi^{k+1}\| = (\sigma + \lambda_{k+1} \|\psi^{k+1}\|) q(x^k, x^{k+1}) \end{aligned}$$

Así obtenemos:

$$\|\varphi^{k+1}\| \leq (\sigma + \lambda_{k+1} \|\psi^{k+1}\|) q(x^k, x^{k+1})$$

Ahora, siempre que $\lambda_{k+1} \|\varphi(x^{k+1})\| \leq M$, donde M es una constante, tenemos la desigualdad:

$$\|\varphi^{k+1}\| \leq (\sigma + M) q(x^k, x^{k+1})$$

y como $\epsilon_k \geq 0 \forall k \in \mathbb{N}$ tenemos:

$$\|\varphi^{k+1}\| \leq (\sigma + M) q(x^k, x^{k+1}) + \epsilon_{k+1}$$

Si consideramos la cuasi-distancia como la distancia Euclidiana tenemos que el Algoritmo 4 cumple la condición (4) de nuestro algoritmo abstracto, las condiciones (2), (3) y (4) las obtenemos directamente de (13), (14) y (15).

4.3. Método de división inexacto hacia adelante - hacia atrás

El siguiente algoritmo extiende el algoritmo introducido en Attouch et al. (2013) Sección 5 [3] al considerar el error escalar en la primera condición.

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia semicontinua inferior y acotada inferiormente. Asumimos que es una función estructurada y que se puede dividir como $f = g + h$, donde $h : \mathbb{R}^n \rightarrow \mathbb{R}$ es de clase C^1 cuya gradiente ∇h es L -Lipschitz continua. Este tipo de problema estructurado ocurre con frecuencia (ver [4] para problemas de sistemas de inclusiones monótonas acopladas en espacios de Hilbert y [9] para problemas de recuperación de señales).

Algoritmo 5

Paso 1 Elegir constantes positivas a, b con $a > L$. Fijar un punto inicial $x^0 \in \text{dom}(g)$ y una sucesión de números reales positivos $\{\epsilon_{k+1}\}$ tal que:

$$\sum_{k=0}^{\infty} \epsilon_{k+1} < \infty \quad (18)$$

Paso 2 Para $k = 0, 1, 2, \dots$ dado $x^k \in \mathbb{R}^n$, encontrar $x^{k+1} \in \mathbb{R}^n$, $v^{k+1} \in \mathbb{R}^n$ tal que:

$$g(x^{k+1}) + \langle x^{k+1} - x^k, \nabla h(x^k) \rangle + \frac{a}{2} \|x^{k+1} - x^k\|^2 \leq g(x^k) + \epsilon_{k+1} \quad (19)$$

$$v^{k+1} \in \partial_{\epsilon_{k+1}} g(x^{k+1}) \quad (20)$$

$$\|v^{k+1} + \nabla h(x^k)\| \leq b \|x^{k+1} - x^k\| \quad (21)$$

Paso 3 Si $x^k = x^{k+1}$ o $0 \in \partial f(x^{k+1})$, parar. Caso contrario, regresar al paso 2.

Vamos a considerar sucesiones generadas por el siguiente algoritmo:

Aplicamos el Lema 2.7 a la función h haciendo $u = x^{k+1}$ y $x = x^k$:

$$h(x^{k+1}) \leq h(x^k) + \langle \nabla h(x^k), x^{k+1} - x^k \rangle + \frac{L}{2} \|x^{k+1} - x^k\|^2$$

al estar en el espacio euclidiano tenemos:

$$h(x^{k+1}) \leq h(x^k) + \langle x^{k+1} - x^k, \nabla h(x^k) \rangle + \frac{L}{2} \|x^{k+1} - x^k\|^2$$

Sumamos en ambos miembros $g(x^{k+1})$ y $\frac{a}{2} \|x^{k+1} - x^k\|^2$ y por (19):

$$\begin{aligned} & h(x^{k+1}) + g(x^{k+1}) + \frac{a}{2} \|x^{k+1} - x^k\|^2 \\ & \leq h(x^k) + \langle x^{k+1} - x^k, \nabla h(x^k) \rangle + \frac{L}{2} \|x^{k+1} - x^k\|^2 + g(x^{k+1}) + \frac{a}{2} \|x^{k+1} - x^k\|^2 \\ & \leq g(x^k) + \epsilon_{k+1} + h(x^k) + \frac{L}{2} \|x^{k+1} - x^k\|^2 \end{aligned}$$

Ordenando, tenemos:

$$h(x^{k+1}) + g(x^{k+1}) + \left(\frac{a-L}{2} \right) \|x^{k+1} - x^k\|^2 \leq g(x^k) + h(x^k) + \epsilon_{k+1}$$

Como $f = h + g$, entonces:

$$f(x^{k+1}) + \left(\frac{a-L}{2} \right) \|x^{k+1} - x^k\|^2 \leq f(x^k) + \epsilon_{k+1}$$

Ahora, definimos $w^{k+1} = v^{k+1} + \nabla h(x^{k+1})$. Veamos que

$$w^{k+1} \in \partial_{\epsilon_{k+1}} f(x^{k+1}) \quad (22)$$

Del algoritmo conocemos que $v^{k+1} \in \partial_{\epsilon_{k+1}} g(x^{k+1})$; luego, por definición:

$$\exists x^l \longrightarrow x^{k+1}, g(x^l) \longrightarrow g(x^{k+1}), v^l \in \hat{\partial}_{\epsilon_{k+1}} g(x^l) \text{ y } v^l \longrightarrow v^{k+1}$$

Como h es continua, tenemos:

$$\exists x^l \longrightarrow x^{k+1} \text{ tal que } f(x^l) = g(x^l) + h(x^l) \longrightarrow g(x^{k+1}) + h(x^{k+1}) = f(x^{k+1}) \quad (23)$$

Por la Proposición 2.4 ítem 2. se cumple que

$$\hat{\partial}_{\epsilon_{k+1}}g(x^l) + \hat{\partial}_0h(x^l) \subset \hat{\partial}_{\epsilon_{k+1}}(g+h)(x^l) = \hat{\partial}_{\epsilon_{k+1}}f(x^l) \quad (24)$$

Luego, por la Proposición 2.3 tenemos $\nabla h(x^l) \in \hat{\partial}_0h(x^l) = \hat{\partial}h(x^l)$ además $v^l \in \hat{\partial}_{\epsilon_{k+1}}g(x^l)$, entonces de (24) obtenemos:

$$v^l + \nabla h(x^l) \in \hat{\partial}_{\epsilon_{k+1}}f(x^l) \quad (25)$$

Como ∇h es continua y $v^l \longrightarrow v^{k+1}$:

$$v^l + \nabla h(x^l) \longrightarrow v^{k+1} + \nabla h(x^{k+1}) = w^{k+1} \quad (26)$$

De (23), (25) y (26) concluimos que:

$$w^{k+1} \in \partial_{\epsilon_{k+1}}f(x^{k+1})$$

Por la propiedad (21) del algoritmo 5, y la desigualdad triangular obtenemos la condición (5) de nuestro algoritmo, como sigue:

$$\begin{aligned} \|w^{k+1}\| &\leq \|v^{k+1} + \nabla h(x^k)\| + \|\nabla h(x^{k+1}) - \nabla h(x^k)\| \\ &\leq b\|x^{k+1} - x^k\| + L\|x^{k+1} - x^k\| \\ &= (b + L)\|x^{k+1} - x^k\| \end{aligned}$$

Luego, como $\epsilon_{k+1} \geq 0$ tenemos:

$$\|w^{k+1}\| \leq (b + L)\|x^{k+1} - x^k\| + \epsilon_{k+1}$$

4.4. Un método de Gauss-Seidel regularizado inexacto

El siguiente algoritmo extiende el algoritmo introducido en Attouch et al. (2013) Sección 6 [3] al considerar el error escalar en la primera condición.

Fijamos un número entero $p \geq 2$, y sean n_1, \dots, n_p enteros positivos. El vector x pertenece al espacio producto $\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$, es denotado por $x = (x_1, \dots, x_p)$, donde cada x_i pertenece a \mathbb{R}^{n_i} . Nos interesa la minimización de funciones $f : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p} \longrightarrow \mathbb{R} \cup \{+\infty\}$ que tienen la siguiente estructura:

$$f(x) = Q(x_1, \dots, x_p) + \sum_{i=1}^p f_i(x_i) \quad (27)$$

donde Q es una función de clase C^1 con gradiente continua localmente Lipschitz, y $f_i : \mathbb{R}^{n_i} \longrightarrow \mathbb{R} \cup \{+\infty\}$ es una función propia semicontinua inferior.

Para cada $i \in \{1, \dots, p\}$, vamos a considerar sucesiones acotadas de matrices simétricas definidas positivas $\{B_i^k\}_{k \in \mathbb{N}} \in \mathbb{R}^{n_i \times n_i}$. Vamos a asumir que los autovalores de las matrices $\{B_i^k : k \in \mathbb{N}, i \in \{1, \dots, p\}\}$ están acotados desde 0; es decir, si b es un autovalor, entonces existe $m > 0$ tal que $|b| > m$.

Algoritmo 6

Paso 1 Tomar $0 < \underline{\lambda} < \bar{\lambda} < \infty$. Para cada $i \in \{1, \dots, p\}$, tomar una sucesión acotada de matrices simétricas definidas positivas $\{A_i^k\}_{k \in \mathbb{N}} \in \mathbb{R}^{n_i \times n_i}$ tal que los autovalores de cada A_i^k ($k \in \mathbb{N}, i \in \{1, \dots, p\}$) están en $[\underline{\lambda}, \bar{\lambda}]$. Tomar parámetros positivos b_i ($i = 1, \dots, p$). Tomar un punto de inicio $x^0 = (x_1^0, \dots, x_p^0) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$. Tomar una sucesión de números reales positivos $\{\epsilon_{k+1}\}$ tal que:

$$\sum_{k=0}^{\infty} \epsilon_{k+1} < \infty \quad (28)$$

Paso 2 Para $k = 0, \dots, p$ encontrar x^{k+1} y $v^{k+1} \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$ tal que:

$$\begin{aligned} f_i(x_i^{k+1}) + Q(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, \dots, x_p^k) + \frac{1}{2} \langle A_i^k(x_i^{k+1} - x_i^k), x_i^{k+1} - x_i^k \rangle \\ \leq f_i(x_i^k) + Q(x_i^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_p^k) + \epsilon_{k+1} \end{aligned} \quad (29)$$

$$v_i^{k+1} \in \partial_{\epsilon_{k+1}} f_i(x_i^{k+1}) \quad (30)$$

$$\|v_i^{k+1} + \nabla_{x_i} Q(x_1^{k+1}, \dots, x_i^{k+1}, x_{i+1}^k, \dots, x_p^k)\| \leq b_i \|x_i^{k+1} - x_i^k\| \quad (31)$$

donde i varía sobre $\{1, \dots, p\}$

Paso 3 Si $x^k = x^{k+1}$ o $0 \in \partial f(x^{k+1})$, parar. Caso contrario, regresar al paso 2.

Definimos $w^{k+1} = (v_i^{k+1} + \nabla_{x_i} Q(x_1^{k+1}, \dots, x_p^{k+1}))_{i=1, \dots, p} \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$ y afirmamos lo siguiente:

$$w^{k+1} \in \partial_{\epsilon_{k+1}} f(x^{k+1}) \quad (32)$$

En efecto, de (30) sabemos que para cada i existe $v_i^{k+1} \in \partial_{\epsilon_{k+1}} f_i(x_i^{k+1})$; entonces, usando las reglas de diferenciación para funciones separadas, tenemos:

$$v^{k+1} = (v_1^{k+1}, \dots, v_p^{k+1}) \in \partial_{\epsilon_{k+1}} \sum_{i=1}^p f_i(x_i^{k+1}) \quad (33)$$

y como Q es de clase C^1 , gracias la demostración de (22) en la subsección anterior, haciendo $h(x^{k+1}) = Q(x^{k+1})$, $g(x^{k+1}) = \sum_{i=1}^p f_i(x_i^{k+1})$, obtenemos que:

$$\nabla Q(x^{k+1}) + v^{k+1} \in \partial_{\epsilon_{k+1}} f(x^{k+1})$$

y así demostramos (32).

Asumiendo que la sucesión $\{x^k\}_{k \in \mathbb{N}}$ es acotado, denotamos por L la constante de Lipschitz de ∇Q sobre el producto de bolas $\bar{B}_1 \times \dots \times \bar{B}_p$ que contienen la sucesión $\{x^k\}_{k \in \mathbb{N}}$. Para todo $i = 1, \dots, p$, tenemos:

$$\begin{aligned} & \|v_i^{k+1} + \nabla_{x_1} Q(x_1^{k+1}, \dots, x_p^{k+1})\| \\ & \leq \|v_i^{k+1} + \nabla_{x_1} Q(x_1^{k+1}, \dots, x_i^{k+1}, x_{i+1}^{k+1}, \dots, x_p^k)\| \\ & \quad + \|Q(x_1^{k+1}, \dots, x_i^{k+1}, x_{i+1}^{k+1}, \dots, x_p^k) - \nabla_{x_1} Q(x_1^{k+1}, \dots, x_p^{k+1})\| \\ & \leq b_i \|x_i^{k+1} - x_i^k\| + L \|x^{k+1} - x^k\| \end{aligned}$$

Por lo tanto, como $\epsilon_{k+1} \geq 0$, para algún $M > 0$ se cumple:

$$\|w^{k+1}\| \leq M \|x^{k+1} - x^k\| + \epsilon_{k+1} \quad (34)$$

Hasta ahora, se cumplen las condiciones (2), (4) y (5) del algoritmo 1. Vamos a demostrar que también se cumple la condición (3), para ello usamos la desigualdad $\lambda \|u\|^2 \leq \langle A_i^k u, u \rangle$ para los enteros i y k en (30) y obtenemos:

$$\begin{aligned} & f_i(x_i^{k+1}) + Q(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, \dots, x_p^k) + \frac{\lambda}{2} \|x_i^{k+1} - x_i^k\|^2 \\ & \leq f_i(x_i^{k+1}) + Q(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, \dots, x_p^k) + \frac{1}{2} \langle A_i^k (x_i^{k+1} - x_i^k), x_i^{k+1} - x_i^k \rangle \\ & \leq f_i(x_i^k) + Q(x_i^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_p^k) + \epsilon_{k+1} \end{aligned}$$

Considerando solo los extremos:

$$\begin{aligned} & f_i(x_i^{k+1}) + Q(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, \dots, x_p^k) + \frac{\lambda}{2} \|x_i^{k+1} - x_i^k\|^2 \\ & \leq f_i(x_i^k) + Q(x_i^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_p^k) + \epsilon_{k+1} \end{aligned} \tag{35}$$

Sumando las desigualdades de (35) desde $i = 1$ hasta $i = p$, concluimos que:

$$f(x^{k+1}) + \lambda \|x^{k+1} - x^k\|^2 \leq f(x^k) + \epsilon_{k+1}$$

De esta manera demostramos que este algoritmo cumple con las condiciones de nuestro algoritmo abstracto 1.

4.5. Ejemplo

Aplicaremos el algoritmo hacia adelante y hacia atrás en el siguiente problema de minimización ambientado en \mathbb{R}^2 .

$$\min \{ \|(x, y)\|_0 + x^2 + y^2 : (x, y) \in \mathbb{R}^2 \}$$

donde $\|\cdot\|_0$ es la norma ℓ^0 ; es decir, para $x \in \mathbb{R}^n$, $\|x\|_0$ es la número de componentes diferentes de 0 de x .

Si hacemos $h(x, y) = x^2 + y^2$ y $g(x, y) = \|(x, y)\|_0$, no es complicado demostrar que cumplen con las condiciones para aplicar el algoritmo.

- $g + h$ es propia y acotada inferiormente pues $\text{dom}(f + g) \neq \emptyset$ y para todo $x \in \mathbb{R}^2$ tenemos que $-\infty < 0 \leq f(x)$.
- $g + h$ es semicontinua inferior, pues h es continua y para g es sencillo demostrar que es semicontinua inferior.
- $\nabla h(x, y) = (2x, 2y)$, luego se cumple que para todo $(x, y), (w, z) \in \mathbb{R}^2$:

$$\|\nabla h(x, y) - \nabla h(w, z)\| = \|(2x, 2y) - (2w, 2z)\| = 2\|(x, y) - (w, z)\|$$

así ∇h es continua y 2-lipschitz, por tanto $L = 2$.

Con las hipótesis demostradas, procedemos a aplicar el algoritmo.

- Tenemos $L = 2$, elegimos $a=6$, $b=8$, $x^0 = (1, 0)$ y $\epsilon_{k+1} = \frac{1}{2^{k+1}}$

- Para $k = 0$, hallamos $x^1 = (x, y)$ y v^1 tal que:

$$g(x, y) + \langle (x, y) - (1, 0), \nabla h(1, 0) \rangle + 3\|(x, y) - (1, 0)\|^2 \leq g(1, 0) + \epsilon_1$$

$$g(x, y) + \langle (x - 1, y), (2, 0) \rangle + 3\|(x - 1, y)\|^2 \leq 1 + \frac{1}{2}$$

$$g(x, y) + 2x - 2 + 3(x - 1)^2 + 3y^2 \leq 1,5$$

Tomamos entonces $x^1 = (0,25, 0)$, dado que al reemplazarlo en la desigualdad anterior resulta $-0,3125 \leq 0$.

Sabemos que $\hat{\partial}_{0,5}g(0,25, 0) := \{v \in \mathbb{R}^2 : \liminf_{y \neq x, y \rightarrow x} \frac{g(y) - g(0,25,0) - \langle v, y - (0,25,0) \rangle}{\|y - (0,25,0)\|} \geq -0,5\}$

Veamos que $v^1 = (0,5, 0) \in \hat{\partial}_{0,5}g(0,25, 0)$

En efecto, llamaremos $G(y) = \frac{g(y) - 1 - \langle (0,5,0), y - (0,25,0) \rangle}{\|y - (0,25,0)\|}$ y analizaremos haciendo $y = (y_1, y_2)$

Notamos que cuando $y_1 \neq 0$ y $y_2 \neq 0 \implies g(y) = 2y$

$$G(y) = \frac{2 - 1 - \langle (0,5,0), y - (0,25,0) \rangle}{\|y - (0,25,0)\|} = \frac{1,125 - 0,5y_1}{\sqrt{(y_1 - 0,25)^2 + y_2^2}} \geq -0,5$$

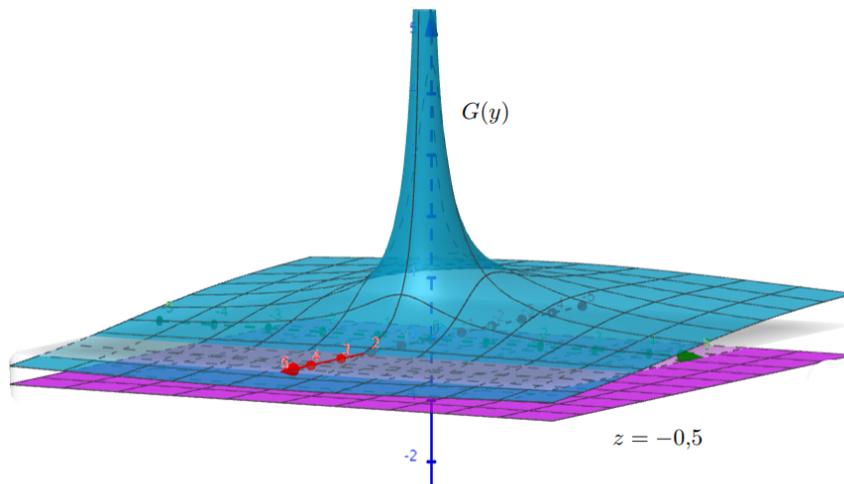


Figura 1: Gráfico de $G(y)$ y el plano $z = -0,5$

Luego, para $r < 0,5$, $\inf \{G(y) : \|x - y\| < r\} \geq -0,5$, entonces, sin necesidad de calcular los para $r \geq 0,5$ inferimos infimos por definición de supremo que:

$$\liminf_{y \neq x, y \rightarrow x} G(y) = \sup_{r > 0} \inf_{y \rightarrow x} \{G(y) : \|x - y\| < r\} \geq -0,5$$

Así $v^1 = (0,5, 0) \in \hat{\partial}_{0,5}g(0,25, 0) \subset \partial_{0,5}g(0,25, 0)$

Y como $\|v^1 + \nabla h(x^0)\| = \|(0,5, 0) + (2, 0)\| \leq 8\|x^0 - x^1\| = \|(1, 0) - (0,25, 0)\|$.

Se concluye que $x^1 = (0,25, 0)$ y $v^1 = (0,5, 0)$

5. Análisis de convergencia

En adelante analizaremos la convergencia de los algoritmos que satisfacen las condiciones de nuestro algoritmo abstracto cuando la función bajo estudio sea propia y semicontinua inferior.

Lema 5.1. (Ver [19], Lema 2 página 44) Sean $\{v_k\}$, $\{\gamma_k\}$ y $\{\beta_k\}$ sucesiones no negativas de números reales que satisfacen $v_{k+1} \leq (1 + \gamma_k)v_k + \beta_k$ y tal que $\sum_{k=1}^{\infty} \beta_k < \infty$, $\sum_{k=1}^{\infty} \gamma_k < \infty$. Entonces la sucesión $\{v_k\}$ converge.

Proposición 5.2. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia extendida semicontinua inferior. La sucesión $\{x^k\}$ generada por algún algoritmo que cumpla las condiciones del algoritmo abstracto, satisface:

(i) $\{f(x^k)\}$ converge.

(ii) $\|x^{k+1} - x^k\|$ converge a cero.

(iii) Si $x^{k_j} \rightarrow \bar{x}$ entonces $x^{k_j+1} \rightarrow \bar{x}$.

Demostración. (i) De (3) tenemos lo siguiente:

$$f(x^{k+1}) \leq f(x^k) - a\|x^{k+1} - x^k\| + \epsilon_{k+1} \implies f(x^{k+1}) \leq f(x^k) + \epsilon_{k+1} \quad (36)$$

entonces:

$$f(x^{k+1}) - \inf f(x) \leq f(x^k) - \inf f(x) + \epsilon_{k+1} \quad (37)$$

Haciendo, para todo $k \in \mathbb{N}$, $g(x^k) = f(x^k) - \inf f(x)$, tenemos:

$$g(x^{k+1}) \leq g(x^k) + \epsilon_{k+1} \implies g(x^{k+1}) \leq (1 + 0)g(x^k) + \epsilon_{k+1} \quad (38)$$

Como $\inf f(x) \leq f(x) \forall x \in \text{dom}(f) \implies 0 \leq f(x) - \inf f(x), \forall x \in \text{dom}(f)$, en particular se cumple para todo $k \in \mathbb{N}$.

De esta manera se cumplen las hipótesis del Lema 5.1; $g(x^k) \geq 0$, (2) y (38), así concluimos que $\{g(x^k)\}$ converge.

Por tanto $\{f(x^k)\}$ converge.

(ii) De (3) tenemos:

$$\|x^{k+1} - x^k\| \leq \frac{1}{a}(f(x^k) - f(x^{k+1})) + \epsilon_{k+1} \quad (39)$$

Como $\{f(x^k)\}$ converge por el ítem (i) y $\{\epsilon_{k+1}\}$ converge a 0 por (2), entonces cuando $k \rightarrow \infty$ tenemos:

$$\frac{1}{a}(f(x^k) - f(x^{k+1})) + \epsilon_{k+1} \rightarrow 0$$

Luego, por (39):

$$0 \leq \lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| \leq 0$$

Así $\|x^{k+1} - x^k\|$ converge a cero.

(iii) De la desigualdad triangular obtenemos:

$$\|x^{k_j+1} - \bar{x}\| \leq \|x^{k_j+1} - x^{k_j}\| + \|x^{k_j} - \bar{x}\|$$

Como $x^{k_j} \rightarrow \bar{x}$ entonces $\|x^{k_j} - \bar{x}\| \rightarrow 0$,

y gracias a (ii) sabemos que $\|x^{k_j+1} - x^{k_j}\| \rightarrow 0$.

Luego, cuando $k \rightarrow \infty$, tenemos que $\|x^{k_j+1} - \bar{x}\| \rightarrow 0$.

Por tanto: $x^{k_j+1} \rightarrow \bar{x}$

□

Observación 5.1. Cuando la función es coerciva, se puede demostrar que la sucesión $\{x^k\}$ es acotada. En efecto, afirmamos que se cumple lo siguiente:

$$f(x^k) \leq f(x^0) + \sum_{i=0}^{k-1} \epsilon_{i+1} \leq f(x^0) + \sum_{i=0}^{\infty} \epsilon_{i+1} < +\infty \quad (40)$$

Actuaremos por inducción sobre k en (40) de la siguiente manera:

- Para $k = 1$, lo obtenemos de (36) cuando $k = 0$, es decir, $f(x^1) \leq f(x^0) + \epsilon_1$
- Suponemos que se cumple para k .
- Veamos que se cumple para $k + 1$:

De la hipótesis inductiva y de (36), tenemos:

$$f(x^{k+1}) - \epsilon_{k+1} \leq f(x^k) \leq f(x^0) + \sum_{i=0}^{k-1} \epsilon_{i+1}$$

esto implica

$$f(x^{k+1}) \leq f(x^0) + \sum_{i=0}^{k-1} \epsilon_{i+1} + \epsilon_{k+1} = f(x^0) + \sum_{i=0}^k \epsilon_{i+1}$$

De (40) haciendo $\alpha_0 = f(x^0) + \sum_{i=0}^{\infty} \epsilon_{i+1}$, tenemos que $x^k \in L_f(\alpha_0) = \{x \in \mathbb{R}^n : f(x) \leq \alpha_0\}$. Como f es coerciva, por la Proposición 2.1 entonces $L_f(\alpha_0)$ es acotada, y por tanto la sucesión $\{x^k\}$ es acotada.

Definimos el siguiente conjunto:

$$U = \{x \in \mathbb{R}^n : f(x) \leq \lim_{k \rightarrow +\infty} f(x^k)\}$$

Proposición 5.3. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia extendida semicontinua inferior. Cada punto de acumulación, si existen, de la sucesión $\{x^k\}$, generada por algún algoritmo que cumpla las condiciones del algoritmo abstracto, pertenece a U .

Demostración. Supongamos que existe $\bar{x} \in \mathbb{R}^n$ un punto de acumulación de $\{x^k\}$, entonces existe $\{x^{k_j}\}$ tal que $x^{k_j} \rightarrow \bar{x}$.

Como f es semicontinua inferior y $\{f(x^k)\}$ converge por la Proposición 5.2 ítem (i), tenemos

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

Luego $\bar{x} \in U$. □

Proposición 5.4. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia extendida semicontinua inferior y continua en su dominio. Cada punto de acumulación, si existen, de la sucesión $\{x^k\}$, generada por algún algoritmo que cumpla las condiciones del algoritmo abstracto, es un punto crítico límite generalizado de f .

Demostración. Supongamos que existe $\bar{x} \in \mathbb{R}^n$ un punto de acumulación de $\{x^k\}$, luego existe una subsucesión $\{x^{k_j}\}$ tal que $x^{k_j} \rightarrow \bar{x}$. Gracias a la Proposición 5.2 ítem (iii) tenemos que $x^{k_j+1} \rightarrow \bar{x}$ y además, $f(x^{k_j+1}) \rightarrow f(\bar{x})$ pues f es continua en su dominio.

De (4) y (5) tenemos:

$$w^{k_j+1} \in \partial_{\epsilon_{k_j+1}} f(x^{k_j+1}) \text{ y } \|w^{k_j+1}\| \leq b\|x^{k_j+1} - x^{k_j}\| + \epsilon_{k+1}$$

Por la Proposición 5.2 ítem (ii) sabemos que $\|x^{k_j+1} - x^{k_j}\| \rightarrow 0$, y gracias a (2) tenemos que $\epsilon_{k+1} \rightarrow 0$, entonces

$$\|w^{k_j+1}\| \rightarrow 0$$

Así tenemos que existe $x^{k_j+1} \rightarrow \bar{x}$ con $f(x^{k_j+1}) \rightarrow f(\bar{x})$, y $w^{k_j+1} \in \partial_{\epsilon_{k_j+1}} f(x^{k_j+1})$ con $w^{k_j+1} \rightarrow 0$.

Luego $0 \in \bar{\partial}f(\bar{x})$. □

6. Algoritmo Abstracto de ϵ -Descenso Generalizado para funciones localmente Lipschitz

Usaremos el ϵ -subdiferencial de Clarke, introducido por primera vez en la sección 5 de Papa et. al. (2022) [18].

Definición 6.1. Para cada $x \in \text{dom}(f)$, el ϵ -subdiferencial de Clarke de f en x denotado por $\partial_\epsilon^\circ f(x)$, es el siguiente conjunto de vectores.

$$\partial_\epsilon^\circ f(x) = \{w \in \mathbb{R}^n \mid f^\circ(x, v) \geq \langle w, v \rangle - \epsilon, \forall v \in \mathbb{R}^n\} \quad (41)$$

donde

$$f^\circ(x, v) = \limsup_{t \downarrow 0, y \rightarrow x} \frac{f(y + tv) - f(y)}{t}$$

A continuación presentamos una variante de nuestro algoritmo abstracto para el caso en que la función objetivo es localmente Lipschitz.

Algoritmo 7 Algoritmo abstracto de ϵ -descenso generalizado para funciones localmente Lipschitz

Paso 1 Elegir $x^0 \in \mathbb{R}^n$, $a, b \in \mathbb{R}$ constantes positivas y una sucesión no negativa $\{\epsilon_{k+1}\} \subset \mathbb{R}$ tal que:

$$\sum_{x=0}^{\infty} \epsilon_{k+1} < \infty \quad (42)$$

Paso 2 Para $k = 0, 1, 2, \dots$ a partir de x^k calculamos x^{k+1} tal que:

$$f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \leq f(x^k) + \epsilon_{k+1} \quad (43)$$

$$w^{k+1} \in \partial_{\epsilon_{k+1}}^\circ f(x^{k+1}) \quad (44)$$

donde

$$\|w^{k+1}\| \leq b\|x^{k+1} - x^k\| + \epsilon_{k+1} \quad (45)$$

Paso 3 Si $x^k = x^{k+1}$ o $0 \in \partial f(x^{k+1})$, parar. Caso contrario, regresar al paso 2.

Proposición 6.1. *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia extendida semicontinua inferior y localmente Lipschitz. Cada punto de acumulación, si existen, de la sucesión $\{x^k\}$, generada por algún algoritmo que cumpla las condiciones del algoritmo abstracto de esta sección, es un punto crítico de Clarke de f .*

Demostración. Supongamos que existe $\bar{x} \in \mathbb{R}^n$ un punto de acumulación de $\{x^k\}$, entonces existe una subsucesión $\{x^{k_j}\}$ tal que $x^{k_j} \rightarrow \bar{x}$. Por la Proposición 5.2 ítem (iii), $x^{k_j+1} \rightarrow \bar{x}$.

De (44), $w^{k_j+1} \in \partial_{\epsilon_{k_j+1}}^\circ f(x^{k_j+1})$, entonces:

$$f^\circ(x^{k_j+1}, v) \geq \langle w^{k_j+1}, v \rangle - \epsilon_{k_j+1} \quad \forall v \in \mathbb{R}^n \quad (46)$$

Por la desigualdad de Cauchy-Schwarz y (45):

$$\langle w^{k_j+1}, v \rangle \leq \|w^{k_j+1}\| \|v\| \leq b \|x^{k_j+1} - x^{k_j}\| \|v\| + \epsilon_{k_j+1} \|v\|$$

Análogo al razonamiento usado en la demostración de la Proposición 5.4, se deduce que:

$$\|w^{k_j+1}\| \|v\| \rightarrow 0 \quad \forall v \in \mathbb{R}^n$$

Por lo tanto, para todo $v \in \mathbb{R}^n$, $\langle w^{k_j+1}, v \rangle \rightarrow 0$.

Tomando \limsup en la desigualdad (46) y usando la propiedad de que $f^\circ(\cdot, \cdot)$ es semicontinua superior obtenemos que: $f^\circ(\bar{x}, v) \geq \limsup_{j \rightarrow +\infty} f^\circ(x^{k_j+1}, v) \geq 0$.

Esto implica que $0 \in \partial^\circ f(\bar{x})$

□

7. Algoritmo Abstracto de ϵ -Descenso Generalizado para funciones convexas

Usaremos el ϵ -subdiferencial de Fenchel, cuya definición es la siguiente.

Definición 7.1. Para cada $x \in \text{dom}(f)$, el ϵ -subdiferencial de Fenchel de f en x , denotado por $\partial_\epsilon^F f(x)$, es el conjunto de vectores $s \in \mathbb{R}^n$ tal que:

$$f(y) \geq f(x) + \langle s, y - x \rangle - \epsilon \quad \text{para todo } y \in \mathbb{R}^n$$

A continuación presentamos una variante de nuestro algoritmo abstracto para el caso convexo.

Proposición 7.1. *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia extendida semicontinua inferior, coerciva y convexa. Cada punto de acumulación, de la sucesión $\{x^k\}$, generada por algún algoritmo que cumpla las condiciones del algoritmo abstracto, es un punto mínimo de f .*

Demostración. Como la función es coerciva, gracias a la Observación 5.1 la función es acotada, y por tanto existe $\bar{x} \in \mathbb{R}^n$ un punto de acumulación de $\{x^k\}$ entonces: $\exists \{x^{k_j}\}$ tal que $x^{k_j} \rightarrow \bar{x}$.

Como f es semicontinua inferior y por la Proposición 5.2 ítem (i), $\{f(x^k)\}$ converge, tenemos que $f(\bar{x}) \leq \lim_{k \rightarrow \infty} f(x^k)$.

Por otro lado, como $w^{k+1} \in \partial_{\epsilon_{k+1}}^F f(x^{k+1})$, por la Definición 7.1 se cumple:

$$f(y) \geq f(x^{k+1}) + \langle w^{k+1}, y - x^{k+1} \rangle - \epsilon_{k+1} \quad \forall y \in \mathbb{R}^n$$

Por la condición (47) del algoritmo sabemos que $\epsilon_{k+1} \rightarrow 0$.

Algoritmo 8 Algoritmo abstracto de ϵ -descenso generalizado para funciones convexas

Paso 1 Elegir $x^0 \in \mathbb{R}^n$, $a, b \in \mathbb{R}$ constantes positivas y una sucesión no negativa $\{\epsilon_{k+1}\} \subset \mathbb{R}$ tal que:

$$\sum_{i \in \mathbb{N}} \epsilon_{i+1} < \infty \quad (47)$$

Paso 2 Para $k = 0, 1, 2, \dots$ a partir de x^k calculamos x^{k+1} tal que:

$$f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \leq f(x^k) + \epsilon_{k+1} \quad (48)$$

$$w^{k+1} \in \partial_{\epsilon_{k+1}}^F f(x^{k+1}) \quad (49)$$

donde

$$\|w^{k+1}\| \leq b\|x^{k+1} - x^k\| + \epsilon_{k+1} \quad (50)$$

Paso 3 Si $x^k = x^{k+1}$ o $0 \in \partial f(x^{k+1})$, parar. Caso contrario, regresar al paso 2.

Además, gracias a las condiciones (47), (50), la Proposición 5.2 ítem (ii) y que $\{x^k\}$ es acotada, tenemos que $\|w^{k+1}\| \rightarrow 0$. Luego, $\langle w^{k+1}, y - x^{k+1} \rangle \rightarrow 0$.

Así, obtenemos que:

$$f(\bar{x}) \leq \lim_{k \rightarrow \infty} f(x^k) \leq f(y) \quad \forall y \in \mathbb{R}^n$$

Por lo tanto, $\bar{x} \in \mathbb{R}^n$ es punto mínimo de f .

□

8. Conclusiones

Presentamos un algoritmo abstracto que abarca los algoritmos inexactos de descenso para encontrar puntos críticos de una función extendida propia y semicontinua inferior, que toma en consideración errores escalares cometidos en cada iteración. Demostramos que cualquier punto de acumulación, si existe, de sucesiones generadas por algoritmos que cumplan con las cuatro condiciones de nuestro algoritmo abstracto, son puntos críticos límite generalizados.

Nos damos cuenta que la coercitividad implica que los conjuntos de niveles inferiores de una función están acotados y, por tanto, la sucesión generada por cualquier algoritmo de descenso está acotada, de este modo, se garantiza la existencia de puntos de acumulación. Lo cual implica cuando la función además es coerciva se puede aplicar directamente las proposiciones de convergencia demostrados en la sección anterior.

En este artículo no utilizamos la propiedad de Kurdyka-Lojaiewicz utilizada en Attouch et al. (2013) [3] para obtener los resultados débiles de convergencia. Sin embargo, valdría la pena estudiar la convergencia global del algoritmo bajo las condiciones de Kurdyka-Lojasiecz, como lo estudió Attouch et al. (2013) [3]. Un estudio sobre esta posibilidad está siendo realizado en el working paper Castillo Ventura y Papa Quiroz [8].

Referencias bibliográficas

- [1] Absil, P. A., Mahony, R. y Andrews, B. (2005). Convergence of the iterates of descent methods for analytic cost functions. *SIAM Journal on Optimization*, Vol. 16, 2: 531 – 547.

- [2] Attouch, H., and Bolte, J., Redont, P. y Soubeyran, A. (2010). Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka–Lojasiewicz inequality. *Mathematics of Operations Research*, Vol. 35, 438 - 457.
- [3] Attouch, H., Bolte, J. y Svaiter, B. (2013). Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, Vol. 137, 1-2: 91 - 129.
- [4] Attouch, H., Briceño-Arias, L. M. y Combettes, P. L. (2010). A parallel splitting method for coupled monotone inclusions. *SIAM Journal on Control and Optimization*, Vol. 48, 5: 3246 – 3270.
- [5] Attouch, H. y Soubeyran, A. (2011). Local Search Proximal Algorithms as Decision Dynamics with Costs to Move. *Set-Valued and Variational Analysis*, Vol. 19, 157 – 177.
- [6] Bao, T. Q., Mordukhovich, B.S., Soubeyran, A. y Tammer, C. (2022). Vector Optimization with Domination Structures: Variational Principles and Applications. *Set-Valued and Variational Analysis*, Vol. 30, 695 - 729.
- [7] Bento, G. C., Ferreira, O. P., Soubeyran, A. y Sousa, J. V. L. (2018). Inexact Multi-Objective Local Search Proximal Algorithms: Application to Group Dynamic and Distributive Justice Problems. *Journal of Optimization Theory and Applications*, Vol. 177, 1: 181 - 200.
- [8] Castillo Ventura E. y Papa Quiroz E. A. (2022) Abstract algorithm of generalized epsilon descent: global convergence and rate of convergence. Working Paper.
- [9] Combettes, P. L. y Wajs, V. R. (2005). Signal recovery by proximal forward–backward splitting. *Multiscale Modeling and Simulation*, Vol. 4, 1168 – 1200.
- [10] Curry, H. B. (2013). The method of steepest descent for non-linear minimization problems. *The Quarterly of Applied Mathematics*, Vol. 1944, 2: 258 – 261.
- [11] de Carvalho, B. G., Bitar, S. D. B., da Cruz, N. J. X., Soubeyran, A. y de Oliveira, S. J. C. (2020). A proximal point method for difference of convex functions in multi-objective optimization with application to group dynamic problems. *Computational Optimization and Applications*, Vol. 75, 1: 263-290.
- [12] Frankel, P., Garrigos, G., Peypouquet J. (2015). Splitting Methods with Variable Metric for Kurdyka - Lojasiewicz Functions and General Convergence Rates. *Journal of Optimization Theory and Applications*, Vol. 165, 3: 874 - 900.
- [13] Hu, Y., Yang, X. y Sim, C. K. (2015). Inexact subgradient methods for quasi-convex optimization problems. *European Journal of Operational Research*.
- [14] Jofre, A., Luc, D. T. y Thera, M. (1998). ϵ -subdiferential and ϵ -monotonicity. *Nonlinear Analysis, Theory, Methods and Applications*, Vol. 33, 1: 71 - 90.
- [15] Palis, J. y Melo, W. (1982). *Geometric theory of dynamical systems*. Springer, New York, Berlin.
- [16] Editor Papa Quiroz, E. A. (2019). *Optimización matemática y computacional*. Papa Quiroz, Erik Alex, ISBN 978-612-00-4468-1.

- [17] Papa Quiroz, E. A., Ramirez, M. y Oliveira P.R. (2015). An inexact proximal method for quasiconvex minimization. *European Journal of Operational Research*, Vol. 246, 721 - 729.
- [18] Papa Quiroz, E. A., Soubeyran, A. y Oliveira, P. R. (2022). Coercivity and generalized proximal algorithms: application—traveling around the world. *Annals of Operations Research*.
- [19] Polyak, B. T. (1987). *Introduction to optimization*. Optimization Software, New York.
- [20] Rockafellar, R. T. y Wets, R. J. B. (1998). *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften. Springer Berlin, Heidelberg, Vol. 317.
- [21] Soubeyran, A. y Souza, J. C. D. O. (2020). General descent method using w-distance. application to emergence of habits following worthwhile moves. *Journal of Nonlinear and Variational Analysis*, Vol. 4, 2: 285 - 300.
- [22] Sun, T., Tang, K. y Li, D. (2022). Gradient Descent Learning with Floats. *IEEE Transactions on Cybernetics*, Vol. 52, 1763 - 1771.