



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Profesional de Matemática

Algoritmo abstracto de ϵ descenso generalizado

TESIS

Para optar el Título Profesional de Licenciada en Matemática

AUTOR

Estéfany CASTILLO VENTURA

ASESOR

Dr. Erik Alex PAPA QUIROZ

Lima, Perú

2024



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Castillo, E. (2024). *Algoritmo abstracto de ϵ descenso generalizado*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Profesional de Matemáticas]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Estéfany Castillo Ventura
Tipo de documento de identidad	DNI
Número de documento de identidad	73027533
URL de ORCID	https://orcid.org/my-orcid?orcid=0000-0001-6502-2356
Datos de asesor	
Nombres y apellidos	Erik Alex Papa Quiroz
Tipo de documento de identidad	DNI
Número de documento de identidad	10451642
URL de ORCID	https://orcid.org/0000-0002-8678-6918
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Alex Armando Cruz Huallpara
Tipo de documento	DNI
Número de documento de identidad	40949706
Miembro del jurado 1	
Nombres y apellidos	Edison Raul Montoro Alegre
Tipo de documento	DNI
Número de documento de identidad	09627181
Datos de investigación	
Línea de investigación	OPTIMIZACIÓN NO CONVEXA Y NO SUAVE

Grupo de investigación	OPTIMACO
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	PERÚ
Año o rango de años en que se realizó la investigación	2022-2023
URL de disciplinas OCDE	MATEMATICA PURA https://purl.org/pe-repo/ocde/ford#1.01.00



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

(Universidad del Perú, DECANA DE AMÉRICA)

FACULTAD DE CIENCIAS MATEMÁTICAS

Ciudad Universitaria - Av. Venezuela S/N cuadra 34

Teléfono: 619-7000, Anexo 1610

Correo Postal: 05-0021, E-mail: eapmat@unmsm.edu.pe

Lima - Perú

Escuela Profesional de Matemática

**ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE
LICENCIADO EN MATEMÁTICA
(Modalidad Presencial)**

En la UNMSM - Ciudad Universitaria - Facultad de Ciencias Matemáticas, siendo las **12:10** horas del martes 05 de marzo del 2024, se reunieron los docentes designados como Miembros del Jurado Evaluador de Tesis: Dr. Edinson Raul Montoro Alegre (PRESIDENTE), Dr. Alex Armando Cruz Huallpara (MIEMBRO) y el Dr. Erik Alex Papa Quiroz (MIEMBRO ASESOR), para la sustentación de la tesis titulada: "**ALGORITMO ABSTRACTO DE ÉPSILON DESCENSO GENERALIZADO**" presentado por la señorita Bachiller **Estéfany Castillo Ventura**, para optar el Título Profesional de Licenciado en Matemática.

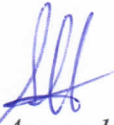
Luego de la exposición de la tesis, el Presidente invitó a la expositora a dar respuesta a las preguntas formuladas.

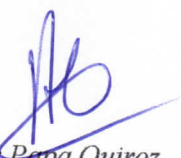
Realizada la evaluación correspondiente por los miembros del Jurado Evaluador, el expositor mereció la aprobación **Sobresaliente** con un calificativo promedio de **diecinueve** (19).

A continuación, el Presidente del Jurado, Dr. Edinson Raul Montoro Alegre, manifestó que la señorita Bachiller **Estéfany Castillo Ventura**, en vista de haber aprobado la sustentación de su tesis, será propuesto para que se le otorgue el Título Profesional de Licenciada en Matemática.

Siendo las **13:18** horas se levantó la sesión firmando para constancia la presente Acta en tres (3) copias originales.


Dr. Edinson Raul Montoro Alegre
PRESIDENTE


Dr. Alex Armando Cruz Huallpara
MIEMBRO


Dr. Erik Alex Papa Quiroz
MIEMBRO ASESOR



CERTIFICADO DE SIMILITUD

Yo Dr. Erik Alex Papa Quiroz en mi condición de asesor acreditado con la Resolución Decanal N°000144-2024-D-FCM/UNMSM de la tesis/monografía/informe de investigación/trabajo académico, cuyo título es **“ALGORITMO ABSTRACTO DE ÉPSILON DESCENSO GENERALIZADO”**, presentado por la bachiller Estéfany Castillo Ventura, para optar el título Profesional de Licenciado en Matemática CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante el software de similitud textual, el documento evaluado cuenta con el porcentaje de 18% de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional.**

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del grado/ título/ especialidad correspondiente.

Firma _____

DNI: 10451642

ERIK ALEX PAPA QUIROZ



Resumen

Dado el problema de minimizar una función posiblemente no convexa y no suave en un espacio real de Hilbert, presentamos un algoritmo de ϵ -descenso generalizado motivado por el método de descenso abstracto introducido por Attouch et al. [22] con dos adiciones esenciales, consideramos errores escalares en la condición de descenso suficiente, así como, en la condición de optimalidad inexacta relativa. Bajo condiciones generales sobre la función a minimizar, obtenemos que todos los puntos de acumulación de las sucesiones generadas por el algoritmo, si existen, son puntos límite críticos generalizados de la función objetivo. Y bajo las condiciones de la desigualdad de ϵ -Kurdyka Lojasiewicz y precompacidad de la función objetivo demostramos convergencia a un punto crítico.

Palabras clave: Métodos de descenso - Optimización no suave y no convexa - Velocidad de convergencia - Algoritmo abstracto.

Abstract

Given the problem of minimizing a possibly nonconvex and nonsmooth function in a real Hilbert space, we present a generalized ε -descent algorithm motivated from the abstract descent method introduced by Attouch et al.[22] with two essential addition, we consider scalar errors on the sufficient descent condition, as well as, on the relative inexact optimality condition. Under general conditions on the function to be minimized, we obtain that all the accumulation points of the sequences generated by the algorithm, if they exist, are generalized critical limit points of the objective function. Moreover, under the ε -Kurdyka Lojasiewicz inequality and if the objective function is precompact, we obtain convergence to critical point.

Keywords: Descent methods - Non-smooth and non-convex optimization - Speed of convergence - Abstract algorithm.

Dedicatoria

Esta tesis va dedicada a mis padres y hermano, por su apoyo incondicional y su fe en mí.

Agradecimientos

En primer lugar, agradezco a Dios Todopoderoso por permitir el cumplimiento de las condiciones necesarias para poder desarrollarme y seguir creciendo en cada aspecto de mi vida, en este caso especial, el aspecto profesional.

Gracias infinitas a mis padres y hermano: Juan Manuel Castillo Pizarro, Betty Ventura Grande y Sebastián Manuel Castillo Ventura; por su amor y apoyo incondicional. Sin ustedes, todo esto no hubiera alcanzado a ver la luz. Su amor, sacrificio y sobre todo su fe en mi han sido el pilar de todo lo logrado.

Me gustaría expresar mi más profundo agradecimiento al Dr. Erik Alex Papa Quiroz por guiarme en este complejo pero gratificante camino de la investigación. Su experiencia, comprensión y paciencia me ayudaron a perseverar y desarrollar esta tesis y de la misma manera a abrir mi mente hacia nuevos horizontes. Su guía constante y su fe inquebrantable en mis habilidades me han motivado a seguir desarrollandome de maneras que nunca imaginé. No tengo palabras para expresar mi gratitud por su inmenso apoyo durante este arduo camino.

Son muchos los docentes que han contribuido en mi formación universitaria, transmitiendo los conocimientos necesarios para hoy estar aquí, A todos ellos quiero agradecer, gracias a ustedes, la confianza en mis conocimientos ha crecido y seguirá creciendo.

Por último agradecer a mi alma mater, que desde un inicio ha sido un reto por exigirme tanto, pero al mismo tiempo me ha permitido obtener mi tan ansiada profesión. Agradezco a cada integrante por su trabajo y por su gestión, sin lo cual no estarían las bases ni las condiciones para aprender.

Tabla de contenidos

Resumen	I
Abstract	II
Dedicatoria	III
Agradecimientos	IV
Tabla de contenidos	V
Introducción	1
I. Marco teórico	4
1.1. Algunas definiciones y resultados del análisis funcional	4
1.2. Notación de Landau - Big O	12
1.3. Épsilon subdiferenciales de Fréchet	13
1.3.1. Subdiferencial de Fréchet y subdiferencial límite de Fréchet	14
1.3.2. ε -subdiferencial de Fréchet y ε -subdiferencial límite de Fréchet	19
1.4. Propiedad ε -Kurdyka Lojasiewicz	23
1.4.1. Desigualdad de Kurdyka Lojasiewicz	23
1.4.2. Desigualdad de ε -Kurdyka Lojasiewicz	29
II. Algoritmo abstracto de ε-descenso generalizado	35
2.1. Algoritmo abstracto de ε -descenso generalizado	35
2.2. Algoritmos con propiedad Epsilon-Descenso	36
2.2.1. Método de gradiente inexacto con errores	36
2.2.2. Método proximal inexacto	40
2.2.3. Algoritmo inexacto de división hacia adelante y atrás con errores	45
2.2.4. Un método de Gauss-Seidel regularizado inexacto	49
III. Convergencia del algoritmo abstracto de ε-descenso generalizado	55
3.1. Algoritmo abstracto de ε -descenso generalizado para funciones localmente lipschitz	58

3.2. Algoritmo abstracto de ϵ -descenso generalizado para funciones convexas	60
3.3. Análisis de convergencia para funciones con la propiedad de ϵ -Kurdyka Lojasiewicz	63
IV. Velocidad de convergencia bajo las condiciones de la desigualdad ϵ-Kurdyka Lojasiewicz	71
V. Ejemplos computacionales	74
VI. Conclusiones	81

INTRODUCCIÓN

La Optimización Matemática esta bastante avanzada en cuanto al estudio de optimizar funciones convexas, y más aún, si estas son suaves; la motivación para desarrollar este proyecto 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 [5], [21], [6], [43], dinámica de grupos [16], [15]).

Encontrar mínimos de una función, que dependiendo del campo de aplicación puede representar diferentes fenómenos, se trata de elegir el mejor de los puntos de su dominio cuya imagen brinde el mejor resultado posible. Este punto, a pesar de los grandes avances en el desarrollo de algoritmos para encontrarlo, en la mayoría de casos no es posible, pero si que se puede conseguir puntos que son candidatos a serlo, a estos los llamamos puntos críticos de una función.

Los métodos desarrollados en la Optimización Matemática, usualmente generan una sucesión de puntos con intensidad que converjan a un punto crítico de la función objetivo. Si esta sucesión es decreciente, estamos frente a un método de descenso, es decir, genera puntos donde la función obtiene valores cada vez más bajos.

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 [24]); en el marco de las ecuaciones diferenciales ocurren comportamientos similares (Ver [27], [1] para ejemplos). Para sortear tales comportamientos nos vemos obligados a trabajar con funciones que poseen ciertas estructuras, por ejemplo, la propiedad de Kurdyka-Lojasiewicz, la coercividad, entre otras.

Ahora bien, cuando implementamos algoritmos en una computadora, se vuelve necesario considerar errores que pueden atribuirse a los datos de entrada, al propio cómputo 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.

Los problemas de minimización pueden estar definidos en espacios euclidianos, variedades Riemannianas, espacios de Hilbert, etc (Ver para variedades Riemannianas [18] y para espacios de Hilbert [19]). Este último espacio mencionado, involucra problemas de EDP cuando son modelados desde el punto de vista variacional, que practicamente es resolver problemas

de optimización en un espacio de funciones. Por tal motivo, desarrollar algoritmos para encontrar puntos críticos de una función en un espacio de Hilbert, alcanza mayores beneficios pues además que es una extensión de un espacio euclidiano también brinda nuevos resultados en el marco de las Ecuaciones Diferenciales Parciales.

Nos enfrentamos entonces al problema de minimizar una función propia semicontinua inferior no suave y no convexa $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ definida en un espacio de Hilbert H mediante métodos de descenso, considerando errores escalares en cada aproximación.

Teniendo en cuenta el problema al cual nos enfrentamos, para minimizar una función no suave y no convexa definida en un espacio de Hilbert, proponemos un algoritmo abstracto de descenso que considere errores escalares.

Inspirándonos en los trabajos realizados por Attouch et al. [22] y ampliado por Frankel et al. [19]. Generalizamos los algoritmos de descenso condesandolos, como sugieren en los artículos mencionados, en dos condiciones: condición de descenso suficiente y condición de error relativo; y añadiendo en ambas errores escalares, es decir, dada una función $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$, el algoritmo propuesto debe generar en un espacio de Hilbert una sucesión $\{x^k\}_{k \in \mathbb{N}}$ que satisface las siguientes condiciones:

- Para cada $k \in \mathbb{N}$:

$$g(x^{k+1}) + \theta \|x^{k+1} - x^k\|^2 \leq g(x^k) + \varepsilon_{k+1}. \quad (0.1)$$

- Para cada $k \in \mathbb{N}$, existe $v^{k+1} \in \partial_{\varepsilon_{k+1}} g(x^{k+1})$:

$$\|v^{k+1}\| \leq \kappa \|x^{k+1} - x^k\| + \beta_{k+1}. \quad (0.2)$$

donde θ, κ son números reales positivos y $\partial_{\varepsilon_{k+1}} g(x^{k+1})$ denota el ε -subdiferencial límite de Fréchet (para detalles de este ε -subdiferencial, ver [4]). La primera condición (0.1), llamada condición de descenso suficiente de $\{g(x^k)\}$, atribuye una medida de descenso y un criterio de error a la sucesión $\{g(x^k)\}$. La segunda condición (0.2), refleja la condición de error relativo para resolver cada subproblema de minimización de un proceso de minimización exacto o inexacto.

Ahora, ¿por qué un algoritmo Abstracto?, respondemos a esta pregunta en las siguientes líneas.

En primer lugar, al desarrollar un algoritmo abstracto, estamos realizando un estudio general gracias al cual podemos abarcar diferentes algoritmos, los cuales ya se usan en paquetes computacionales para resolver problemas de optimización. Por tal motivo, las propiedades o resultados de convergencia que obtendremos, brindará a su vez nuevos resultados para estos algoritmos. Y de la misma manera, al ser condiciones generales, estos pueden ser tomados en el desarrollo de nuevos métodos.

En segundo lugar, en el análisis numérico existen dos tipos de errores: errores de truncamiento (generado a causa de involucrar procesos de aproximación) y errores de redondeo (generado por reemplazar un número por su representación de punto flotante) (ver [13] para más detalles del análisis de errores en Análisis Numérico). Estos errores afectan en la obtención de

resultados exactos, pues el algoritmo que teóricamente funciona, en la práctica no es tanto así. Por eso se debe analizar algoritmos que tengan errores en sus iteraciones para saber hasta que punto se puede controlar el error. Razón por la cual estamos investigando en un marco general algoritmos de descenso considerando errores escalares en cada aproximación.

Ahora, si bien en Frankel et al. (2015) [19] se extendió el trabajo realizado en Attouch et al. (2013) [22] al desarrollar un algoritmo abstracto que considera un error escalar en la segunda condición (la condición del error relativo), en este trabajo, consideramos errores en las dos condiciones, tanto en la condición de descenso suficiente como en la condición del error relativo, brindando así mayor consistencia en el tratamiento de errores de nuestro algoritmo abstracto de descenso. De esta manera nuestros principales aportes pueden ser resumidos en los siguientes:

- 1 Se desarrolla un algoritmo abstracto de descenso, que involucra errores escalares en cada aproximación, para encontrar puntos críticos de una función no suave y no convexa definida en un espacio de Hilbert.
- 2 Especializamos el algoritmo abstracto a métodos de descenso conocidos como el gradiente, proximal, división hacia adelante y hacia atrás, Gauss-Seidel.
- 3 Presentamos convergencia a un punto crítico generalizado bajo supuestos suaves y convergencia a un punto crítico bajo la desigualdad de Kurdyka Lojasiewicz.

Nuestro trabajo está dividido en seis capítulos más los anexos. El primer capítulo está destinado a los principales conceptos y resultados que se utilizaron durante la investigación, conceptos de análisis funcional, los ϵ -subdiferenciales de Fréchet, sus definiciones y propiedades, basado principalmente en [4], presentamos también la desigualdad de Kurdyka Lojasiewicz y en cómo llegamos a una nueva versión denominada desigualdad de ϵ -Kurdyka Lojasiewicz. En el segundo capítulo presentamos el algoritmo abstracto de ϵ -descenso generalizado y algunas de sus especializaciones: método de gradiente inexacto con errores, dos métodos del tipo proximal con errores, algoritmo inexacto de división hacia adelante y hacia atrás con errores y por último el método inexacto de Gauss Seidel con errores. En el tercer capítulo se encuentran nuestros principales resultados de convergencia y sus especializaciones para funciones más específicas: convexas, lipschitzianas, ϵ -Kurdyka Lojasiewicz. El cuarto capítulo se dedica a presentar un estudio de la velocidad de convergencia bajo la propiedad de ϵ -Kurdyka Lojasiewicz. Nuestro quinto capítulo contiene algunos resultados computacionales que demuestran que estos algoritmos con errores convergen. En nuestro sexto capítulo están nuestras conclusiones y alusiones para futuras investigaciones.

Capítulo I

Marco teórico

1.1. Algunas definiciones y resultados del análisis funcional

Espacios normados y espacios de Banach

Definición 1.1.1 (Definición 2.2-1 página 58 [31]) Sea E un espacio vectorial. Una norma en el espacio E es una función $\|\cdot\| : E \rightarrow \mathbb{R}^+$ que satisface:

$$(i) \quad \|x\| \geq 0 \quad \forall x \in E$$

$$(ii) \quad \|x\| \iff x = 0$$

$$(iii) \quad \|\lambda x\| = |\lambda| \|x\|, \quad \forall \lambda \in \mathbb{R}, \quad \forall x \in E$$

$$(iv) \quad \|x + y\| \leq \|x\| + \|y\|, \quad \forall x, y \in E$$

La norma en E define una métrica d dada por:

$$d(x, y) = \|x - y\| \quad x, y \in E$$

y es llamada métrica inducida por la norma. Un espacio normado es un espacio vectorial E dotado de una norma denotado por $(E, \|\cdot\|)$ o simplemente E .

Definición 1.1.2 (Definición 2.2-1 página 59 [31]) Un espacio normado E es un espacio de Banach si es completo (toda sucesión de Cauchy es convergente) para la distancia inducida por su norma.

Funcionales lineales y continuos. Espacio Dual

Conocemos como operadores a funciones definidas en espacios vectoriales.

Definición 1.1.3 (Operador lineal, Definición 2.6-1 página 82 [31])

Un operador lineal T es un operador tal que

(i) El dominio y el rango del operador son espacios vectoriales sobre el mismo campo.

(ii) Para todo x, y en el dominio y el escalar α ,

$$T(x+y) = T(x) + T(y)$$

$$T(\alpha x) = \alpha T(x)$$

Definición 1.1.4 (Operador lineal acotado, Definición 2-7-1 página 91 [31]) Dados E y V dos espacios normados y $T : \text{dom}(T) \subset E \rightarrow V$ un operador lineal. Decimos que T es acotado si existe $C \geq 0$ tal que

$$\|T(x)\|_V \leq C\|x\|_E, \forall x \in \text{dom}(T)$$

donde $\text{dom}(T)$ es el dominio de T .

Teorema 1.1.1 (Continuidad y acotación, Teorema 2.7-9 página 97 [31]) Sea $T : \text{dom}(T) \rightarrow V$ un operador lineal, con $\text{dom}(T) \subset E$ y E, V dos espacios normados entonces:

(i) T es continua si y solo si T es acotado.

(ii) Si T es continuo en un punto entonces es continuo en todo su dominio.

Definición 1.1.5 (Funcional lineal, Definición 2.8-1 página 104 [31])

Un funcional lineal f es un operador lineal con dominio en un espacio vectorial E y un rango en el campo escalar K de E ; así:

$$f : \text{dom}(f) \longrightarrow K$$

donde $K = \mathbb{R}$ si E es real y $K = \mathbb{C}$ si E es complejo.

Proposición 1.1.2 (Continuidad y acotación, Teorema 2.8-3 página 104 [31]) Un funcional lineal f con dominio $\text{dom}(f)$ en un espacio normado es continuo si y solo si f es acotado.

Ahora, dados E y V dos espacios normados, consideramos el espacio $\mathcal{L}(E, V)$ de todos los operadores lineales acotados definidos en E sobre V , es decir, todo su dominio es E y su rango es V .

Teorema 1.1.3 (Teorema 2.10-1 página 118 [31]) El espacio $\mathcal{L}(E, V)$ es un espacio normado.

Teorema 1.1.4 (Teorema 2.10-2 página 118 [31]) Si V es un espacio de Banach, entonces $\mathcal{L}(E, V)$ es un espacio de Banach.

Definición 1.1.6 (Espacio Dual, Definición 2.10-3 página 119 [31]) El espacio dual de un espacio normado E , se denota como E^* , es el espacio de todos los funcionales lineales acotados, $\mathcal{L}(E, \mathbb{R})$.

Teorema 1.1.5 (Teorema 2.10-4 página 119 [31]) El espacio dual E^* de un espacio normado E es un espacio de Banach.

Producto interno. Espacios de Hilbert

Definición 1.1.7 (*Producto interno, espacio de Hilbert, Definición 3.1-1 página 128 [31]*) El producto interno sobre un espacio vectorial E es una aplicación $\langle \cdot, \cdot \rangle : E \times E \longrightarrow \mathbb{R}$ que satisface:

$$(i) \quad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle, \quad \forall x, y, z \in E.$$

$$(ii) \quad \langle \alpha x, y \rangle = \alpha \langle x, y \rangle, \quad \forall \alpha \in \mathbb{R}, \quad \forall x \in E$$

$$(iii) \quad \langle x, y \rangle = \overline{\langle y, x \rangle}, \quad \forall x, y \in E$$

$$(iv) \quad \langle x, x \rangle \geq 0, \quad \forall x \in E$$

$$(v) \quad \langle x, x \rangle = 0 \iff x = 0.$$

Un espacio vectorial E dotado de un producto interno se conoce como prehilbertiano.

Proposición 1.1.6 (*Espacio de Hilbert, Definición 3.1-1 página 129 [31]*) Un espacio prehilbertiano E es un espacio normado con la norma inducida por el producto interno:

$$\|x\| = \sqrt{\langle x, x \rangle}, \quad \forall x \in E$$

Definición 1.1.8 (*Espacio de Hilbert, Definición 3.1-1 página 129 [31]*) Un espacio prehilbertiano H es Hilbert si es un espacio de Banach con la norma inducida por el producto interno.

Lema 1.1.7 (*Desigualdad de Schwarz, Lema 2.1-1 página 136 [31]*) El producto interno y su norma inducida cumple con la siguiente desigualdad:

$$|\langle x, y \rangle| \leq \|x\| \|y\|, \quad \forall x, y \in E$$

donde E es un espacio prehilbertiano.

Teorema 1.1.8 (*Teorema de Riesz, Teorema 3.8-1 página 188 [31]*) *Todo funcional lineal f en un espacio de Hilbert puede representarse en términos del producto interno:*

$$f(x) = \langle x, z \rangle$$

donde z depende de f , se determina unívocamente por f y su norma:

$$\|z\| = \|f\|$$

Es necesario especificar que a partir de este momento, nos centraremos en los espacios de Hilbert denotados por H , de modo que las definiciones y resultados que vamos a presentar están ligados a las características de este espacio. En caso de presentar la definición en otro espacio se especificará.

Definición 1.1.9 (*Operador Hilbert-adjunto T^* , Definición 3.10-1 página 201 [31]*) *Sea $T : H_1 \rightarrow H_2$ un operador lineal acotado, donde H_1 y H_2 son espacios de Hilbert. Entonces el operador Hilbert-adjunto T^* de T es el operador:*

$$T^* : H_2 \rightarrow H_1$$

tal que para todo $x \in H_1$ y $y \in H_2$,

$$\langle Tx, y \rangle = \langle x, T^*y \rangle$$

Definición 1.1.10 (*Operador autoadjunto, Definición 3.9-1 página 196 [31]*) *Un operador lineal acotado $T : H \rightarrow H$ sobre un espacio de Hilbert H se dice autoadjunto o Hermitiano si*

$T^* = T$. La definición anterior se convierte:

$$\langle Tx, y \rangle = \langle x, Ty \rangle$$

Convergencia fuerte y débil

Definición 1.1.11 (Sucesión de Cauchy, Sección 2.3. página 63 [31]) Una sucesión $\{x_n\}$ en un espacio normado X es Cauchy si para todo $\varepsilon > 0$ existe un N tal que: $\|x_m - x_n\| < \varepsilon \forall m, n \in N$

Definición 1.1.12 (Convergencia fuerte, Definición 4.8-1 página 256 [31]) Una sucesión $\{x_n\}$ en H es fuertemente convergente (o convergente en la norma) si H contiene x tal que:

$$\lim_{n \rightarrow \infty} \|x_n - x\| = 0$$

Entonces escribimos $x_n \rightarrow x$ y llamamos a x el límite de $\{x_n\}$.

Definición 1.1.13 (Convergencia débil, Definición 4.8-2 página 257 [31])

Una sucesión $\{x_n\}$ en un espacio normado E se dice débilmente convergente si existe $x \in E$ tal que para todo $f \in E^*$

$$\lim_{n \rightarrow \infty} f(x_n) = f(x)$$

esto se escribe

$$x_n \rightharpoonup \bar{x}$$

Observación 1.1.1 (Definición 3.2. página 57 [11]). La noción de convergencia débil viene de una topología, denotada por $\sigma(E, E^*)$, que es la topología más gruesa sobre E y que hace continua a todas las aplicaciones $(\phi_f)_{f \in E^*}$, donde cada $\phi_f : E \rightarrow \mathbb{R}$ es un funcional lineal por $\phi_f(x) = \langle f, x \rangle$. Y como $\sigma(E, E^*)$ es una topología menos fina que la topología definida por la norma tiene menos abiertos y menos cerrados; luego, en particular, si un conjunto es cerrado

para esta topología entonces es cerrado para la topología inicial de E y en general la recíproca es falsa.

Observación 1.1.2 (Ejemplo 4.8-5 página 260 [31]) En un espacio de Hilbert $x^k \rightarrow \bar{x}$ si y solo si $\langle x_n, z \rangle \rightarrow \langle \bar{x}, z \rangle, \forall z \in H$.

Proposición 1.1.9 (Sección 4.8. problema 3, página 261 [31]). Si $\{x^k\}, \{y^k\} \subset H$ converge débilmente a \bar{x} y \bar{y} respectivamente, entonces $\{x^k + y^k\}$ converge débilmente a $\bar{x} + \bar{y}$.

Función semicontinua inferior fuerte y función semicontinua inferior débil

Definición 1.1.14 (Definición 1.4. página 10 [11]). Sea V un espacio topológico definido por la norma en V y $g : V \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia. Decimos que g es semicontinua inferior sobre V si para todo $\beta \in \mathbb{R}$ el siguiente conjunto es cerrado.

$$\{x \in V : g(x) \leq \beta\}.$$

Observación 1.1.3 ((Definición 1.4. ítem 2 página 10 [11])) Tomando $V = H$, como H es un espacio métrico, la definición anterior es equivalente a: Si $\{x^k\} \subset H$ converge fuertemente a algún $\bar{x} \in H$, entonces $\liminf_{k \rightarrow \infty} g(x^k) \geq g(\bar{x})$.

Definición 1.1.15 (Corolario 3.9. página 61 [11]). Sea $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ una función semicontinua inferior respecto a la topología débil, esto es, el conjunto $\{x \in V : g(x) \leq \beta\}$ es cerrado en la topología débil, entonces decimos que la función g es débilmente semicontinua inferior

Observación 1.1.4 (Observación 3 página 60 [11]) Ya que la topología débil no es metrizable, si g es una función semicontinua inferior entonces, si $\{x^k\} \subset H$ converge débilmente a algún $x \in H$ tenemos

$$\liminf_{k \rightarrow \infty} g(x^k) \geq g(x),$$

lo contrario no es cierto en general.

Proposición 1.1.10 (Corolario 3.9. observación 6, página 61 [11]). Dado una función semicontinua inferior $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$. Si g es también convexa entonces g es débilmente semicontinua inferior.

Proposición 1.1.11 (Teorema 3.18, página 69 [11]). Sea $\{x^k\} \subset H$ una sucesión acotada, entonces existe un punto de acumulación débil, i.e., existe $\{x^{k_j}\} \subset \{x^k\}$ una subsucesión débilmente convergente.

Definición 1.1.16 (Función coerciva, página 6 [20]). Una función $g : H \rightarrow \mathbb{R}$ es coerciva sobre H si

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

Proposición 1.1.12 (Teorema 3.2.5. página 87 [30]) Si $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ es una función propia, coerciva y débilmente semicontinua inferior; entonces existe algún $u \in H$ tal que $f(u) \leq f(v)$, $\forall v \in H$.

Lema 1.1.13 ([38] Lema 9.2 página 64) Si $g : H \rightarrow \mathbb{R}$ es una función de clase C^1 con gradiente L -Lipschitz, entonces la siguiente desigualdad se cumple para todo x, y en H :

$$g(y) \leq g(x) + \langle \nabla g(x), y - x \rangle + \frac{L}{2} \|y - x\|^2.$$

Definición 1.1.17 Sea la función $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, decimos que la sucesión $\{x^k\}$ es g -precompacta si y solo si tiene una subsucesión x^{k_j} que cumple $x^{k_j} \rightarrow x^*$ y $g(x^{k_j}) \rightarrow g(x^*)$.

Definición 1.1.18 El conjunto ℓ^1 es definido como sigue:

$$\ell^1 = \left\{ \{x^l\}_{l \in \mathbb{N}} : \sum_{l=0}^{+\infty} |x^l| < +\infty \right\}. \quad (1.1)$$

1.2. Notación de Landau - Big O

La notación de Landau ilustra da origen a los conocidos "*O* mayúscula" (Big-*O*) y "*o* minúscula" (little-*o*), el utilizado en este trabajo será el primero, que ilustraremos a continuación

Definición 1.2.1 (*Definición 1.1.3. página 6 [37]*) Sean f y g dos funciones definidas en un subconjunto de \mathbb{R}^n .

(1) $f(x) = O(g(x))$ cuando $x \rightarrow x_0$ si y solo si $\exists M, \delta \in \mathbb{R}_+$ tal que $|f(x)| \leq M|g(x)|$ para $|x - x_0| < \delta$. Cuando $g(x_0) \neq 0$ se puede presentar de otra manera:

$$\limsup_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} < \infty$$

(2) $f(x) = O(g(x))$ cuando $x \rightarrow \infty$ si y solo si $\exists M, y \in \mathbb{R}_+$ tal que $|f(x)| \leq M|g(x)|$ para todo $x \geq y$

Proposición 1.2.1 $O(f(n)) + O(g(n)) = O(\max(f(n), g(n)))$

Demostración:

Sin perder generalidad suponemos que $f(n) \leq g(n)$, luego

$$O(f(n)) + O(g(n)) = c_1 \cdot f(n) + c_2 \cdot g(n)$$

Así que podemos escribir

$$O(f(n)) + O(g(n)) \leq c_1 \cdot g(n) + c_2 \cdot g(n) \leq (c_1 + c_2)g(n) \leq c \cdot g(n) \leq c \cdot \max(f(n), g(n))$$

Por la definición:

$$O(f(n)) + O(g(n)) = O(\max(f(n), g(n)))$$



1.3. Épsilon subdiferenciales de Fréchet

En Optimización convexa, el concepto de subgradiente aproximado o ε -subgradiente se ha convertido en una útil herramienta, para cuyo desarrollo se ha contado con numerosos e importantes resultados. De modo que los ε -subgradientes desempeñan un papel central en la construcción de métodos numéricos para la optimización de funciones convexas (Ver el libro de Hiriart-Urruty y Lemaréchal [25] Cap. XI, Vol. II).

Para funciones no convexas también se ha desarrollado este concepto mediante diferentes enfoques. Por ejemplo: en [26] los autores definen el concepto de ε -subdiferenciabilidad local; en [3] se introdujo el concepto para funciones localmente lipschitz, un ε -subgradiente cuyo rango es el casco convexo de los valores existentes de la gradiente de la función y sus límites en una ε -bola cerrada; en [45] aparece el concepto de gradiente-bola para una clase de funciones diferenciables casi en todas partes sobre \mathbb{R}^n , en [39] para funciones localmente Lipschitz sobre \mathbb{R}^n , el concepto ε -delocalizar gradiente; en [2] se definió la noción de $\varepsilon - \eta$ -gradiente; en [14] conceptos de gradiente generalizado ε -local y gradiente generalizado ε -global.

Ahora bien, para funciones generales extendidas de valor real y sobre la base de límites sucesiónles de las llamadas Fréchet ε -normales, en [46] introdujeron un concepto de subdiferencial que goza de un rico cálculo y permite derivar condiciones necesarias y suficientes para la apertura, la regularidad métrica, el comportamiento lipschitziano de las multifunciones con aplicaciones a la optimización, el análisis de sensibilidad. (Revisar [12]).

En este trabajo utilizaremos el llamado ε -subdiferencial límite de Fréchet; sin embargo comenzaremos esta sección presentado definiciones y algunos resultados para los subdiferenciales de Fréchet.

1.3.1. Subdiferencial de Fréchet y subdiferencial límite de Fréchet

Los subdiferenciales de Fréchet se conocen desde hace más de un cuarto de siglo. Probablemente fueron introducidas por primera vez en dimensiones finitas en [9] (bajo el nombre de semidiferenciales inferiores"). Antes de ir a por nuestro objetivo, mencionaremos al subdiferencial de Fenchel.

Definición 1.3.1 (Subdiferencial de Fenchel, 2.1 página 72 [4]) El subdiferencial de Fenchel de g en $x \in \text{dom}(g)$, denotado por $\partial^F g(x)$, es definido por el conjunto:

$$\partial^F g(x) = \{v \in H : g(z) \geq g(x) + \langle v, z - x \rangle, \forall z \in H\}.$$

Definición 1.3.2 (Subdiferencial de Fréchet, Definición 1.1 página 3326 [32])

Sea $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ y un punto x en su dominio, entonces el conjunto:

$$\hat{\partial}g(x) := \left\{ w \in H : \liminf_{\substack{z \rightarrow x \\ z \neq x}} \frac{g(z) - g(x) - \langle w, z - x \rangle}{\|z - x\|} \geq 0 \right\}$$

es llamado Subdiferencial de Fréchet.

Observación 1.3.1 Cuando $\hat{\partial}g(x) \neq \emptyset$ diremos que g es Fréchet subdiferenciable.

Definición 1.3.3 (Subdiferencial límite de Fréchet, Definición 1.9 página 3335 [32])

Sea $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ una función semicontinua inferior en una vecindad de x , el conjunto

$$\partial g(x) := \{w \in H : \exists x^l \rightarrow x, g(x^l) \rightarrow g(x), w^l \in \hat{\partial}g(x^l) \text{ con } w^l \rightarrow w\}$$

es llamado Subdiferencial límite de Fréchet.

Observación 1.3.2 Dada una función $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ y un punto $\bar{x} \in \text{dom}(g)$, entonces se cumple que $\hat{\partial}g(\bar{x}) \subset \partial g(\bar{x})$.

Proposición 1.3.1 (Proposition 1.1., página 3326 [32]) Dada una función propia $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ y Fréchet diferenciable en \bar{x} , con derivada $Dg(\bar{x})$, obtenemos que $\hat{\partial}g(\bar{x}) = \{Dg(\bar{x})\}$.

Proposición 1.3.2 (Ejercicio 8.8, página 304 [42]) Dada una función propia $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ diferenciable en una vecindad de \bar{x} , obtenemos que $\partial g(\bar{x}) = \{\nabla g(\bar{x})\}$.

Proposición 1.3.3 (Proposition 1.7., página 3326 [32]) Si $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ es convexa, entonces:

$$\hat{\partial}g(\bar{x}) = \{w \in H : g(u) - g(v) \geq \langle w, u - v \rangle \forall u, v \in H\}$$

Proposición 1.3.4 (Proposition 1.7., página 3328 [32]) Si $\hat{\partial}g(\bar{x}) \neq \emptyset$, entonces g es semicontinua inferior en \bar{x} .

Proposición 1.3.5 (Proposición 1.10, página 3328 [32]). Dada $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia con un mínimo local, $\bar{x} \in \text{dom}(g)$, entonces $0 \in \hat{\partial}g(\bar{x})$.

Proposición 1.3.6 (Proposición 1.11, página 3328 [32]) $\hat{\partial}g(\lambda\bar{x}) = \lambda\hat{\partial}g(\bar{x})$ para cualquier $\lambda > 0$.

Proposición 1.3.7 (Proposición 1.12, página 3328 [32]) Sea $g_1, g_2 : H \rightarrow \mathbb{R} \cup \{+\infty\}$ dos funciones Fréchet subdiferenciables en \bar{x} . Entonces $g_1 + g_2$ es también Fréchet subdiferenciable en \bar{x} y:

$$\hat{\partial}g_1(\bar{x}) + \hat{\partial}g_2(\bar{x}) \subset \hat{\partial}(g_1 + g_2)(\bar{x})$$

Proposición 1.3.8 (Teorema 2.33. página 216 [12]). Sea $g, h : H \rightarrow \mathbb{R} \cup \{+\infty\}$ funciones Fréchet subdiferenciables en \bar{x} , donde g es una función semicontinua inferior y h localmente Lipschitz, entonces $g + h$ es Fréchet subdiferenciable en \bar{x} :

$$\hat{\partial}g(\bar{x}) + \hat{\partial}h(\bar{x}) \supset \hat{\partial}(g + h)(\bar{x}).$$

Observación 1.3.3 *La Proposición 1.3.8 es verdadera cuando el espacio de partida es un espacio de Asplund, pero como trabajamos en espacios de Hilbert los cuales son espacios de Asplund, usamos libremente la Proposición anterior en todo el documento. (Para demostrar que los espacios de Hilbert son espacios de Asplund basta revisar [17] y [40] dónde demuestran que todo espacio reflexivo de Banach es de Asplund, por ende, un espacio de Hilbert es de Asplund).*

Ejemplo 1.3.1 *Dada la función $g(x) = -|x|$, $\forall x \in \mathbb{R}$, calculamos $\hat{\partial}g(x)$. Por la Proposición 1.3.1 tenemos:*

- *Cuando $x > 0$: $\hat{\partial}g(x) = \{-1\}$.*
- *Cuando $x < 0$: $\hat{\partial}g(x) = \{1\}$.*

Sin embargo, veamos que la función no es Fréchet subdiferenciable en 0, es decir, $\hat{\partial}g(0) = \emptyset$.

Por definición:

$$\hat{\partial}g(0) := \left\{ w \in \mathbb{R} : \liminf_{z \rightarrow 0, z \neq 0} -1 - \frac{wz}{|z|} \geq 0 \right\}$$

Sin embargo, no es difícil notar que $\liminf_{z \rightarrow 0, z \neq 0} -1 - \frac{wz}{|z|} \leq -1 \forall w \in \mathbb{R}$. Luego $\hat{\partial}g(0) = \emptyset$.

En efecto:

- *Si $w = 0$, luego $\liminf_{z \rightarrow 0, z \neq 0} -1 = -1$*
- *Si $w < 0$, luego $\liminf_{z \rightarrow 0, z \neq 0} -1 - \frac{wz}{|z|} = -1 + w < -1$*
- *Si $w > 0$, luego $\liminf_{z \rightarrow 0, z \neq 0} -1 - \frac{wz}{|z|} = -1 - w < -1$*

Ejemplo 1.3.2 *Dada la función:*

$$g(x) = \begin{cases} x \operatorname{sen}\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Calculemos el $\hat{\partial}g(x)$

- Cuando $x \neq 0$ tenemos $\hat{\partial}g(x) = \left\{ \text{sen}\left(\frac{1}{x}\right) - \frac{\cos\left(\frac{1}{x}\right)}{x} \right\}$.
- Cuando $x = 0$ tenemos

$$\hat{\partial}g(0) := \left\{ w \in \mathbb{R} : \liminf_{z \rightarrow 0, z \neq 0} \frac{z \text{sen}\left(\frac{1}{z}\right) - wz}{|z|} \geq 0 \right\}$$

Es evidente nuevamente que $\liminf_{z \rightarrow 0, z \neq 0} \frac{z \text{sen}\left(\frac{1}{z}\right) - wz}{|z|} \leq -1$.

En efecto:

- Para $w = 0$, $\liminf_{z \rightarrow 0, z \neq 0} \frac{z \text{sen}\left(\frac{1}{z}\right)}{|z|} = -1$
- Para $w > 0$, $\liminf_{z \rightarrow 0, z \neq 0} \frac{z \text{sen}\left(\frac{1}{z}\right) - wz}{|z|} = -1 - w < -1$
- Para $w < 0$, $\liminf_{z \rightarrow 0, z \neq 0} \frac{z \text{sen}\left(\frac{1}{z}\right) - wz}{|z|} = -1 + w < -1$

Así, nuevamente tenemos $\hat{\partial}g(0) = \emptyset$.

Observación 1.3.4 Sin embargo, si consideramos la función:

$$g(x) = \begin{cases} \text{máx}\{x \text{sen}\left(\frac{1}{x}\right), 0\} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

El inconveniente se presenta en 0, veamos que en este caso $\hat{\partial}g(0) = \{0\}$

- Para $w = 0$, $\liminf_{z \rightarrow 0, z \neq 0} \frac{\text{máx}\{z \text{sen}\left(\frac{1}{z}\right), 0\}}{|z|} = 0$
- Para $w > 0$, $\liminf_{z \rightarrow 0, z \neq 0} \frac{\text{máx}\{z \text{sen}\left(\frac{1}{z}\right), 0\} - wz}{|z|} = -w < 0$
- Para $w < 0$, $\liminf_{z \rightarrow 0, z \neq 0} \frac{\text{máx}\{z \text{sen}\left(\frac{1}{z}\right), 0\} - wz}{|z|} = w < 0$

Ejemplo 1.3.3 Ahora, consideremos la siguiente función:

$$g(x) = \begin{cases} x^2 \operatorname{sen}\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Calculemos $\hat{\partial}g(x)$ y $\partial g(x)$.

Empecemos con $\hat{\partial}g(x)$.

- Cuando $x \neq 0$, $\hat{\partial}g(x) = \{2 \cdot x \cdot \operatorname{sen}\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)\}$
- Cuando $x = 0$, $\hat{\partial}g(0) = \{\hat{\partial}g(0) := \{w \in \mathbb{R} : \liminf_{z \rightarrow x, z \neq x} \frac{z^2 \operatorname{sen}\left(\frac{1}{z}\right) - zw}{|z|} \geq 0\}\}$
 - Para $w = 0$: $\liminf_{z \rightarrow x, z \neq x} \frac{z^2 \operatorname{sen}\left(\frac{1}{z}\right)}{|z|} = 0 \geq 0$
 - Para $w \neq 0$: $\liminf_{z \rightarrow x, z \neq x} \frac{z^2 \operatorname{sen}\left(\frac{1}{z}\right) - wz}{|z|} < 0$

Luego:

$$\hat{\partial}g(x) = \begin{cases} 2 \cdot x \cdot \operatorname{sen}\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Continuamos calculando $\partial g(x)$.

- Cuando $x \neq 0$ tenemos por Proposición 1.3.2, $\partial g(x) = \{2 \cdot x \cdot \operatorname{sen}\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)\}$.
- Cuando $x = 0$: $\partial g(x) = [-1, 1]$. En efecto: Sea $x^k \xrightarrow{g} 0$, luego existen $w^k \in \hat{\partial}g(x^k)$. Luego:

$$\lim_{k \rightarrow \infty} 2 \cdot x^k \cdot \operatorname{sen}\left(\frac{1}{x^k}\right) - \cos\left(\frac{1}{x^k}\right) = \lim_{k \rightarrow \infty} -\cos\left(\frac{1}{x^k}\right)$$

Así: $\partial g(0) = [-1, 1]$

1.3.2. ε -subdiferencial de Fréchet y ε -subdiferencial límite de Fréchet

En esta sección presentaremos las definiciones y algunos resultados de los ε -subdiferenciales de Fréchet, para ello consideraremos a B como la bola unitaria cerrada con centro en cero en H . Y como en la sección anterior comenzaremos mencionando al ε -subdiferencial de Fenchel

Definición 1.3.4 (*Observación 2.4. página 75 [4]*) *El ε -subdiferencial de Fenchel de g en $x \in \text{dom}(g)$, denotado por $\partial_\varepsilon^F g(x)$, esta definido por el siguiente conjunto:*

$$\partial_\varepsilon^F g(x) = \{v \in H : g(z) \geq g(x) + \langle v, z - x \rangle - \varepsilon, \forall z \in H\}.$$

Definición 1.3.5 ((2.3) [4]) *Sea $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ una función, $x \in \text{dom}(g)$, ε un número fijo no negativo, el ε -subdiferencial de Fréchet de g en x es definido por:*

$$\hat{\partial}_\varepsilon g(x) := \left\{ w \in H : \liminf_{\|h\| \rightarrow 0} \frac{g(x+h) - g(x) - \langle w, h \rangle}{\|h\|} \geq -\varepsilon \right\}.$$

Observación 1.3.5 ((2.5) [4]) *Una definición equivalente es: $w \in \hat{\partial}_\varepsilon g(x)$ si y solo si para cada $\eta > 0$, existe $\delta > 0$ tal que para todo $y \in x + \delta B$*

$$\langle w, y - x \rangle \leq g(y) - g(x) + (\varepsilon + \eta)\|y - x\|$$

Observación 1.3.6 ((2.6) [4]) *Si g es convexa, entonces*

$$\hat{\partial}_\varepsilon g(x) = \{w \in H : g(x+h) - g(x) - \langle w, h \rangle \geq -\varepsilon\|h\| \forall h \in H\}$$

Proposición 1.3.9 (*Propiedades del ε -subdiferencial de Fréchet página 73 [4]*). *Sean $h, g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ funciones propias, entonces:*

1. $\hat{\partial}_\varepsilon h(x)$ es vacío si $x \notin \text{dom}(h)$,

2. $\hat{\partial}_\varepsilon h(x)$ es un subconjunto convexo de H .
3. $\hat{\partial}_{\varepsilon_1 + \varepsilon_2}(h + g)(x) \supset \hat{\partial}_{\varepsilon_1} h(x) + \hat{\partial}_{\varepsilon_2} g(x)$,
4. Si $\varepsilon_1 \leq \varepsilon_2$ entonces $\hat{\partial}_{\varepsilon_2} h(x) \supset \hat{\partial}_{\varepsilon_1} h(x)$.
5. Si $\hat{\partial}_\varepsilon h(x)$ es no vacío, entonces h es semicontinua inferior en x .

Definición 1.3.6 ((2.7) y (2.8) [4]) A partir de los datos en la Definición 1.3.5, llamamos ε -subdiferencial límite de Fréchet de g en x al conjunto dado por:

$$\partial_\varepsilon g(x) = \limsup_{y \xrightarrow{g} x} \hat{\partial}_\varepsilon g(y)$$

y de manera equivalente se define:

$$\partial_\varepsilon g(x) := \{w \in H : \exists x^l \rightarrow x, g(x^l) \rightarrow g(x), w^l \in \hat{\partial}_\varepsilon g(x^l) \text{ con } w^l \rightarrow w\}.$$

Proposición 1.3.10 (Proposición 2.3, page 74 [4]) Sea $\varepsilon > 0$ fijado y sea $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ una función convexa. Entonces:

$$\hat{\partial}_\varepsilon g(x) = \partial^F g(x) + \varepsilon B$$

y

$$\partial_\varepsilon g(x) = \hat{\partial}_\varepsilon g(x)$$

Proposición 1.3.11 (Proposición 2.8 página 77 [4]). Sea $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ una función continuamente Fréchet diferenciable en x , con derivada $Dg(x)$. Entonces,

$$\partial_\varepsilon g(x) = Dg(x) + \varepsilon B$$

Definición 1.3.7 (Definición 2.14 página 80 [4]) Sea $x_0 \in \text{dom}(g)$. Decimos que x_0 es un ε -mínimo de g si

$$g(x_0 + h) \geq g(x_0) - \varepsilon, \forall h \in H$$

y se dice que x_0 es un $\varepsilon\|\cdot\|$ -mínimo de g si

$$g(x_0 + h) \geq g(x_0) - \varepsilon\|h\|, \forall h \in H$$

Proposición 1.3.12 (Proposición 2.15 página 80 [4]) Si \bar{x} es un $\varepsilon\|\cdot\|$ -mínimo de g entonces $0 \in \partial_\varepsilon g(\bar{x})$. Si x_0 es un ε -mínimo de g , entonces para todo $\delta > 0$ existe \bar{x} tal que $\|\bar{x} - x_0\| < \delta$ y \bar{x} es un $\varepsilon\|\cdot\|$ -mínimo de g .

Teorema 1.3.13 (Teorema 2.17 página 81 [4]) Sean $f, g : H \rightarrow \mathbb{R}$ dos funciones semicontinuas inferiores y una de ellas localmente lipschitz. Entoces para cada $x \in \text{dom}(f) \cup \text{dom}(g)$ y cada $\varepsilon > 0$ la siguiente inclusión se cumple:

$$\partial_\varepsilon(f + g)(x) \subset \partial_\varepsilon f(x) + \partial_\varepsilon g(x)$$

Definición 1.3.8 (Definición 2.1. (e) página 4 [5]) El subdiferencial límite generalizado de g en x es el conjunto:

$$\bar{\partial}g(x) := \{w \in H : \exists x^l \rightarrow x, g(x^l) \rightarrow g(x), w^l \in \partial_{\varepsilon_l} g(x^l) w^l \rightarrow w \text{ y } \varepsilon_l \rightarrow 0\}.$$

Ejemplo 1.3.4 Dada la función $g(x) = -|x|$, $\forall x \in \mathbb{R}$, calculamos $\hat{\partial}_\varepsilon g(x)$. Por la Proposición 1.3.1 tenemos:

- Cuando $x > 0$: $\hat{\partial}_\varepsilon g(x) = [-1 - \varepsilon, -1 + \varepsilon]$.
- Cuando $x < 0$: $\hat{\partial}_\varepsilon g(x) = [1 - \varepsilon, 1 + \varepsilon]$.

Sin embargo, veamos que $\hat{\partial}_\varepsilon g(0)$ dependerá del valor de ε . Por definición.

$$\hat{\partial}g(0) := \left\{ w \in \mathbb{R} : \liminf_{z \rightarrow 0, z \neq 0} -1 - \frac{wz}{|z|} \geq -\varepsilon \right\}$$

Sin embargo, no es difícil notar que $\liminf_{z \rightarrow 0, z \neq 0} -1 - \frac{wz}{|z|} \leq -1 \forall w \in \mathbb{R}$. Luego $\hat{\partial}g(0) = \mathbf{0}$.

En efecto:

- Si $w = 0$, luego $\liminf_{z \rightarrow 0, z \neq 0} -1 = -1$
- Si $w < 0$, luego $\liminf_{z \rightarrow 0, z \neq 0} -1 - \frac{wz}{|z|} = -1 + w < -1$
- Si $w > 0$, luego $\liminf_{z \rightarrow 0, z \neq 0} -1 - \frac{wz}{|z|} = -1 - w < -1$

Si tomamos $\varepsilon = 1$, luego $\hat{\partial}_1 g(0) = \{0\}$

Ejemplo 1.3.5 Veamos que el ε -subdiferencial límite de Fréchet de la función $h : \mathbb{R} \rightarrow \mathbb{R}$

$h(x) = |x|$ es:

$$\partial_\varepsilon h(x) = \begin{cases} [1 - \varepsilon, 1 + \varepsilon] & \text{si } x > 0 \\ [-1 - \varepsilon, 1 + \varepsilon] & \text{si } x = 0 \\ [-1 - \varepsilon, -1 + \varepsilon] & \text{si } x < 0 \end{cases}$$

En efecto,

- Cuando $x > 0$ la derivada es 1 y por la Proposición 1.3.11 se tiene que $\partial_\varepsilon h(x) = [1 - \varepsilon, 1 + \varepsilon]$, $x > 0$.
- Cuando $x < 0$ la derivada es -1 y por la Proposición 1.3.11 se tiene que $\partial_\varepsilon h(x) = [-1 - \varepsilon, -1 + \varepsilon]$, $x < 0$.

- Cuando $x = 0$ debemos notar que como la función $h(x) = |x|$ es convexa podemos usar las Proposiciones 1.3.10 y 1.3.6. Luego:

$$\partial_\varepsilon h(0) = \hat{\partial}_\varepsilon h(0) = \{w \in H : |z| - wz \geq -\varepsilon|z| \forall z \in H\}$$

- Si $z > 0$, en la desigualdad tenemos $z - wz \geq -\varepsilon z$, luego $w \in (-\infty, 1 + \varepsilon]$
- Si $z < 0$, en la desigualdad tenemos $-z - wz \geq \varepsilon z$, luego $w \in [-1 - \varepsilon, +\infty)$
- Si $z = 0$, entonces w adquiere cualquier valor.

Luego concluimos que $\partial_\varepsilon h(0) = [-1 - \varepsilon, 1 + \varepsilon]$.

1.4. Propiedad ε –Kurdyka Lojasiewicz

1.4.1. Desigualdad de Kurdyka Lojasiewicz

La desigualdad de Kurdyka-Łojasiewicz es un resultado fundamental en análisis matemático y optimización, que proporciona una condición para la convergencia global de los algoritmos de optimización y la estabilidad de soluciones en problemas de optimización no lineales.

La primera versión de la desigualdad fue introducida por Lojasiewicz ([35], [36]) como parte fundamental de la demostración que establecía que para una función analítica convexa f , cualquier trayectoria de descenso de gradiente acotada tiene longitud finita. La desigualdad de gradiente de Lojasiewicz es enunciada de la siguiente manera:

Proposición 1.4.1 (Proposición 1, pagina 67 (92) de [35])

Sea f una función analítica en la vecindad de un punto a de un espacio euclídeo M . Si $f(a) = 0$ entonces existe un $0 < \theta < 1$ tal que:

$$|\nabla f(x)| \geq |f(x)|^\theta$$

en una vecindad de a .

Observación 1.4.1 En la formulación original piden $f(a) = 0$; sin embargo actualmente se la presenta de la siguiente manera:

Proposición 1.4.2 (Lema 2.1. página 533 de [7]) Sea f una función analítica en una vecindad del punto $a \in \mathbb{R}^n$. Entonces existe una constante $c > 0$ y $\mu \in [0, 1)$ tal que

$$\|\nabla f(x)\| \geq c|f(x) - f(a)|^\mu \quad (1.2)$$

en alguna vecindad U de a

Posteriormente fue extendida para funciones subanalíticas, por Kurdyka y Parusinski en la década de 1990 [29] de la siguiente manera:

Proposición 1.4.3 ([29]) Sea $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ una función subanalítica, diferenciable en $\Omega \setminus f^{-1}(0)$, donde Ω es un subconjunto abierto de \mathbb{R}^n . Entonces existe $C > 0$, $\rho > 0$ y $0 \leq \alpha < 1$ tal que:

$$|\nabla f(x)| \geq C|f(x)|^\alpha \quad (1.3)$$

para cada $x \in \Omega$ tal que $|f(x)| \in (0, \rho)$.

Si además $\lim_{x \rightarrow a} f(x) = 0$ para algun $a \in \bar{\Omega}$ entonces la desigualdad (1.3) se cumple para todo $x \in \Omega \setminus f^{-1}(0)$ cercano a a .

Para 1980, Kurdyka extendió la desigualdad para funciones definidas en estructuras o-minimales [33]:

Proposición 1.4.4 (Teorema 1 página 775 [33]) Sea $f : U \rightarrow \mathbb{R}$ una M -función diferenciable donde U es un conjunto abierto de \mathbb{R}^n . Suponemos que $f(x) > 0 \forall x \in U$. Entonces existe $C > 0$, $\rho > 0$ y una M -función estrictamente positiva y creciente $\Psi : \mathbb{R}_+ \rightarrow \mathbb{R}$ de clase C^1 tal que $\|\nabla(\Psi \circ f)(x)\| \geq C \forall x \in U, f(x) \in (0, \rho)$

Alrededor del año 2007, Jérôme Bolte, Aris Daniilidis y Adrian Lewis [10] extendieron la desigualdad de Lojasiewicz a funciones subanalíticas no lisas, estableciendo una desigualdad análoga en la que la derivada $\nabla f(x)$ fue reemplazado por algún elemento w del subdiferencial $\partial f(x)$. Y al igual que su versión suave este resultado proporciona nuevas ideas sobre los aspectos de convergencia de los sistemas dinámicos de tipo subgradiente. Los teoremas más importantes que nos dan en [10] son:

Proposición 1.4.5 (Teorema 3.1. página 1213 [10]) Sea $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ una función subanalítica con dominio cerrado y asumimos que $f|_{\text{dom}(f)}$ es continua. Sea $a \in \mathbb{R}^n$ un punto crítico de f . Entonces existe un exponente $\theta \in [0, 1)$ tal que:

$$\frac{|f(x) - f(a)|^\theta}{m_f(x)} \quad (1.4)$$

es acotado para todo x alrededor de a . Donde

$$m_f(x) = \inf\{\|x^*\| : x^* \in \partial f(x)\}$$

Proposición 1.4.6 (Teorema 3.3. página 1215 [10]) Sea $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ una función semicontinua inferior, convexa y subanalítica con $\text{crit}(f) \neq \emptyset$. Para cualquier conjunto acotado K existe un exponente $\theta \in [0, 1)$ tal que:

$$\frac{|f(x) - f(a)|^\theta}{m_f(x)} \quad (1.5)$$

es acotado en K . Donde

$$m_f(x) = \inf\{\|x^*\| : x^* \in \partial f(x)\}$$

La desigualdad anterior se nos presenta en un nuevo formato en el 2010 de parte de Attouch et al. [8]. Para ello tomaremos en cuenta lo siguiente:

Sea $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ una función propia semicontinua inferior. Para dos constantes $-\infty < \eta_1 < \eta_2 \leq +\infty$, se define el conjunto:

$$[\eta_1 < f < \eta_2] = \{x \in \mathbb{R}^n : \eta_1 < f(x) < \eta_2\}$$

Definición 1.4.1 (Definición 3.2, página 445 [8])

(a) La función $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ se dice que posee la propiedad de Kurdyka Lojasiewicz en $x^* \in \text{dom}(\partial f)$ si existe $\eta \in (0, +\infty]$, una vecindad U de x^* y una función concava continua $\varphi : [0, \eta) \longrightarrow \mathbb{R}_+$ tal que:

(i) $\varphi(0) = 0$

(ii) φ es de clase C^1 sobre $(0, \eta)$

(iii) Para todo $s \in (0, \eta)$, $\varphi'(s) > 0$

(iv) Para todo $x \in U \cap [f(x^*) < f < f(x^*) + \eta]$, la desigualdad de Kurdyka-Lojasiewicz se cumple:

$$\varphi'(f(x) - f(x^*)) \text{dist}(0, \partial f(x)) \geq 1$$

(b) Las funciones propias y semicontinuas inferiores que cumplen con la desigualdad de Kurdyka Lojasiewicz en cada punto del $\text{dom}(\partial f)$ es llamada función KL.

Es aquí donde demuestran que una función $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ propia y semicontinua inferior cumple con la desigualdad KL en cualquier punto no crítico $\hat{x} \in \mathbb{R}^n$, específicamente en la Observación 3.2. (b) de [8].

Durante el mismo año 2010, en [28] nos presentan caracteriaciones de la desigualdad KL en espacios métricos, con especial consideración en los espacios de Hilbert.

Comenzaremos presentando la caracterización en un espacio métrico completo, para ello es necesario presentar unas definiciones extras.

Sean X, Y dos espacios métricos completos, $F : X \rightrightarrows Y$ un mapeo multivaluado y (\hat{x}, \hat{y}) en el gráfico de F .

Definición 1.4.2 (Definición 1, página 3324 [28]) Sea $k \in [0, +\infty)$.

(i) El mapeo multivaluado F es llamado k -métricamente regular en (\hat{x}, \hat{y}) en el gráfico de F si existe $\varepsilon, \delta > 0$ tal que $(x, y) \in B(\hat{x}, \varepsilon) \times B(\hat{y}, \delta)$ tenemos:

$$d(x, F^{-1}(y)) \leq k d(y, F(x))$$

(ii) Sea V un subconjunto no vacío de $X \times Y$. El mapeo multivaluado F es llamado k -métricamente regular sobre V si F es métricamente regular (\hat{x}, \hat{y}) para todo (\hat{x}, \hat{y}) en el gráfico de F intersectado con V .

Proposición 1.4.7 (Corolario 4, página 3327 [28]) Sea $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ una función semicontinua inferior definida en espacio métrico completo X y $\varphi \in \Gamma(0, r_0)$. Asumiendo que $[0 < f < r_0] \neq \emptyset$. Entoces las siguientes afirmaciones son equivalentes:

(i) El mapeo multivaluado:

$$\begin{cases} X & \rightrightarrows & \mathbb{R} \\ x & \mapsto & [(\varphi \circ f)(x), +\infty] \end{cases}$$

es k -métricamente regular en $[0 < f < r_0] \times (0, \varphi(r_0))$;

(ii) Para $r_1, r_2 \in (0, r_0)$,

$$d([f \leq r_1], [f \leq r_2]) \leq k |\varphi(r_1) - \varphi(r_2)|$$

(iii) Para todo $x \in [0 < f < r_0]$,

$$|\nabla(\varphi \circ f)|(x) \geq \frac{1}{k}$$

Si cualquiera de estas aseveraciones se cumple, el conjunto de nivel $[f \leq 0]$ es no vacío.

En la sección 3 del mismo artículo, Teorema 18, nos brindan muchas equivalencias de la desigualdad KL en espacios de Hilbert para funciones no suaves y, a pesar de no ser muy restrictiva, para que el teorema que proponen se cumpla se necesita la condición de semiconvexidad. Sin embargo, en general concluimos que una versión suave de la desigualdad KL en espacios de Hilbert (H), la podemos considerar de la siguiente manera:

Proposición 1.4.8 (Teorema 18, página 3333 [28]) Sea $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ una función semicontinua inferior. Decimos que cumple con la desigualdad de Kurdyka-Lojasiewicz si existe $r_0 \in (0, \bar{r})$, $\varepsilon \in (0, \bar{\varepsilon})$ y $\phi \in \mathbb{K}(0, \bar{r}_0)$ tal que:

$$\|\partial(\phi \circ f)(x)\|_- \geq 1 \quad \text{for all } x \in B(\bar{x}, \varepsilon) \cap [0 < f \leq \bar{r}_0] \quad (1.6)$$

Ejemplo 1.4.1 Veamos que la función $h(x) = |x|$, $x \in \mathbb{R}$ satisface la desigualdad de Kurdyka Lojasiewicz

Sabemos que $h(x) = x$ cuando $x > 0$ y $h(x) = -x$ cuando $x < 0$. Por tanto en ese dominio $h(x)$ es semialgebraica, lo que implica que cumple con la propiedad Kurdyka Lojasiewicz. Solo queda demostrar que la propiedad KL se cumple en $x = 0$.

Consideramos $\varphi(t) = t$, $\eta, \delta > 0$

- $\varphi(0) = 0$
- φ es de C^1 en todo \mathbb{R}
- $\forall s \in \mathbb{R} \quad \varphi'(s) = 1 > 0$

- Sea $x \in B(0, 1) \cap [0 < h(x) < \eta]$, luego

$$\varphi'(h(x)) \text{dist}(0, \partial h(x)) = \|\nabla h(x)\| = 1 \geq 1$$

Ejemplo 1.4.2 En general la desigualdad se cumple para $h(x) = \|x\|^p$ con $x \in \mathbb{R}^n$ donde $0 < \frac{1}{p} < 1$

Notemos que h es no convexa y no suave. Veamos que es una función KL.

Sea $x^* \in \mathbb{R}^n$, elegimos $\varphi(t) = t^{\frac{1}{p}}$, $\eta > 0$ y $\delta > 0$

- φ es de clase C^1 en \mathbb{R}
- $\varphi(0) = 0$
- $\forall t \in (0, \eta) \quad \varphi'(t) = \frac{1}{p} t^{\frac{1}{p}-1} > 0$
- Si $x \in B(x^*, \delta) \cup [0 < h < \eta]$, entonces:

$$\frac{1}{p} (\|x\|^p)^{\frac{1}{p}-1} p \|x\|^{p-1} = 1 \geq 1$$

1.4.2. Desigualdad de ε -Kurdyka Lojasiewicz

En esta sección explicaremos el razonamiento tras la modificación de la desigualdad de Kurdyka Lojasiewicz que utilizamos para el análisis de convergencia del algoritmo propuesto, pues al considerar errores escalares permitimos cierta tolerancia a las inexactitudes en los resultados, por tal motivo, la herramienta que utilizamos también debe ser adaptada para tal objetivo. Además demostramos que la modificación presentada conserva sus propiedades.

Empezamos presentado una modificación de la primera versión de la desigualdad:

Definición 1.4.3 Sea f una función analítica en la vecindad de un punto de un espacio euclidiano M , que verifica $f(a) = 0$ y sea $0 \leq \gamma < \infty$ de modo que para $\varepsilon \in [0, \gamma]$ se cumple la

desigualdad de ε -Lojasiewicz si existe un $0 < \theta < 1$ tal que

$$\text{dist}(\partial_\varepsilon f(x), 0) \geq \|f(x)\|^\theta$$

en una vecindad de a .

Observación 1.4.2 Como nos movemos en los espacios de Hilbert y al ser la función objetivo analítica, por la Proposición 1.3.11 tenemos:

$$\partial_\varepsilon f(x) = \nabla f(x) + \varepsilon B$$

De este modo, para entender el cambio realizado, reemplacemos este valor en la desigualdad anterior:

$$\text{dist}(\hat{\partial}_\varepsilon, 0) = \inf\{\nabla f(x) + \varepsilon B\} \geq \|f(x)\|^\theta$$

Luego para $b \in B$, tenemos

$$\|\nabla f(x)\| + \|\varepsilon b\| \geq \inf\{\nabla f(x) + \varepsilon B\} \geq \|f(x)\|^\theta$$

Tomando $\|b\| = 1$ la desigualdad queda:

$$\|\nabla f(x)\| + \varepsilon \geq \|f(x)\|^\theta$$

Cómo vemos la norma de la gradiente sigue controlando los valores de $f(x)$ pero ahora permitiendo un margen de error y si ese margen de error desaparece nos encontramos con la desigualdad original.

Ahora, considerando la versión que Kurdyka extendió para funciones definidas en estructuras o-minimales en [33].

Teorema 1.4.9 Sea f una M -función diferenciable, donde U es un conjunto acotado y abierto de \mathbb{R}^n . Suponga que $f(x) > 0$ para todo $x \in U$. Se dice que la función cumple con desigualdad de ε - Kurdyka Lojasiewicz si existe $c > 0$, $\rho > 0$, $\gamma \in [0, \infty)$ y una M -función estrictamente creciente positiva $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ de clase C^1 , tal que para $\varepsilon \in [0, \gamma]$

$$\phi'(f(x)) \text{dist}(\partial_\varepsilon f(x), 0) \geq c$$

para cada $x \in U$, $f(x) \in (0, \rho)$.

Observación 1.4.3

$$\phi'(f(x)) \text{dist}(\hat{\partial}_\varepsilon f(x), 0) \geq c$$

Como la función es diferenciable, tenemos gradiente: Elegimos un elemento $y \in \hat{\partial}_\varepsilon f(x)$ entonces $y = \nabla f(x) + \varepsilon b$, para algún $b \in B$ tal que $\|b\| = 1$, luego:

$$\phi'(f(x)) \|y\| \geq c$$

$$\phi'(f(x)) \|\nabla f(x) + \varepsilon b\| \geq c$$

$$\phi'(f(x)) \|\nabla f(x)\| + \phi'(f(x)) \|\varepsilon b\| \geq c$$

$$\|\phi'(f(x)) \nabla f(x)\| + \phi'(f(x)) \varepsilon \geq c$$

Luego la desigualdad queda:

$$\|\nabla(\phi \circ f)(x)\| + \phi'(f(x)) \varepsilon \geq c$$

Y como se puede ver tenemos la desigualdad original pero con un margen de error que es lo que necesitamos.

Ahora, para funciones no convexas y no suaves, en el formato dado en la Definición 1.4.1 tendríamos:

Definición 1.4.4 (*Propiedad de ε -Kurdyka-Lojasiewicz*) La función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ tiene la propiedad de ε -Kurdyka-Lojasiewicz en $x^* \in \text{dom}(\partial f)$ si existe $\eta \in (0, +\infty]$, $\gamma > 0$, una vecindad U en x^* y una función concava continua $\varphi : [0, \eta] \rightarrow \mathbb{R}_+$ tal que:

(i) $\varphi(0) = 0$

(ii) φ is C^1 on $(0, \eta)$

(iii) Para todo $s \in (0, \eta)$, $\varphi'(s) > 0$

(iv) Para todo $x \in U \cap [f(x^*) < f < f(x^*) + \eta]$ y para $\varepsilon \in [0, \gamma]$, la desigualdad de ε -Kurdyka-Lojasiewicz cumple:

$$\varphi'(f(x) - f(x^*)) \text{dist}(0, \partial_\varepsilon f(x)) \geq 1.$$

De esta manera le brindamos mayor flexibilidad a la desigualdad para poder tratar a los algoritmos con errores.

Proposición 1.4.10 Si $0 \notin \partial f(\bar{x})$ entonces existe $c > 0$, tal que para todo x y todo $\varepsilon > 0$ tal que $\|x - \bar{x}\| + |f(x) - f(\bar{x})| + \varepsilon < c$. Entonces:

$$\text{dist}(0, \partial_\varepsilon f(x)) \geq c$$

Demostración:

Por contradicción, supongamos que existe una sucesión $\{c^k\}$ que converge a 0 y una sucesión $\{x^k\}$ y $\{\varepsilon_k\}$ tal que $\|x^k - \bar{x}\| + |f(x^k) - f(\bar{x})| + \varepsilon_k < c^k$ y $\text{dist}(0, \partial_{\varepsilon_k} f(x^k)) < c^k$. Entonces existe $x_c^k \in \partial_{\varepsilon_k}(x^k)$ tal que $\|x_c^k\| < c^k$. Luego tenemos que existe $x^k \rightarrow \bar{x}$, $f(x^k) \rightarrow f(\bar{x})$, $\varepsilon_k \rightarrow 0$ y $x_c^k \rightarrow 0$ con $x_c^k \in \partial_{\varepsilon_k}(x^k)$, lo que implica que $0 \in \partial f(\bar{x})$. ■

Proposición 1.4.11 Si $0 \notin \partial f(\bar{x})$ entonces f satisface la desigualdad de ε - Kurdyka Lojasiewicz en \bar{x} .

Demostración:

De la proposición anterior sabemos que existe $c > 0$ de modo que para $x \in B(\bar{x}, \frac{c}{3}) \cup [f(\bar{x}) < f < f(\bar{x}) + \frac{c}{3}]$ y $\varepsilon \in [0, \frac{c}{3}]$. Considerando $\phi(t) = \frac{1}{c}t$ la cual es una función cóncava, continua y diferenciable, con valores positivos para $x \in [0, \frac{c}{3}]$ y derivada positiva. Luego:

$$\frac{1}{c} \text{dist}(0, \partial_\varepsilon f(x)) \geq 1$$

De este modo se demuestra que la función cumple la desigualdad de ε - Kurdyka Lojasiewicz en todo punto no crítico de la función.



Ejemplo 1.4.3 Veamos que la función $h(x) = |x|$, $x \in \mathbb{R}$ satisface la desigualdad de Kurdyka Lojasiewicz

Sabemos que $h(x) = x$ cuando $x > 0$ y $h(x) = -x$ cuando $x < 0$. Por tanto en ese dominio $h(x)$ es semialgebraica, lo que implica que cumple con la propiedad Kurdyka Lojasiewicz. Solo queda demostrar que la propiedad KL se cumple en $x = 0$. Para ello, gracias al Ejemplo 1.3.5 sabemos:

$$\partial_\varepsilon h(x) = \begin{cases} [1 - \varepsilon, 1 + \varepsilon] & \text{si } x > 0 \\ [-1 - \varepsilon, 1 + \varepsilon] & \text{si } x = 0 \\ [-1 - \varepsilon, -1 + \varepsilon] & \text{si } x < 0 \end{cases}$$

Consideramos $\varphi(t) = 2t$, $\eta, \delta > 0$

- $\varphi(0) = 0$
- φ es de C^1 en todo \mathbb{R}

- $\forall s \in \mathbb{R} \quad \varphi'(s) = 2 > 0$
- Sea $x \in B(0, 1) \cap [0 < h(x) < \eta]$, luego

$$\varphi'(h(x)) \text{dist}(0, \partial_\varepsilon h(x)) = 2|1 - \varepsilon| = |2 - 2\varepsilon|$$

Luego, existe $\gamma \leq 0,5$ para $\varepsilon \in [0, 0,5]$ tal que

$$\varphi'(h(x)) \text{dist}(0, \partial_\varepsilon h(x)) \geq 1$$

Capítulo II

Algoritmo abstracto de ϵ -descenso generalizado

2.1. Algoritmo abstracto de ϵ -descenso generalizado

En esta sección introducimos el algoritmo abstracto de ϵ -descenso generalizado que se inspira en el trabajo de Attouch et al. [22], con la diferencia de que en este trabajo consideramos errores escalares en la condición de descenso suficiente (0.1), un ϵ -subdiferencial y un error sobre la acotación del error relativo en (0.2).

La motivación para introducir este algoritmo inexacto es considerar en los algoritmos de descenso los errores computacionales debidos a aproximaciones, datos de entrada o ruido.

Algoritmo 1 Algoritmo abstracto de épsilon-descenso generalizado

Paso 0: Elegir $x^0 \in H$, dos números reales positivos θ, κ y dos sucesiones reales no negativas $\{\varepsilon_{k+1}\}, \{\beta_{k+1}\} \subset \mathbb{R}$ tal que:

$$\{\varepsilon_{k+1}\}_{k \in \mathbb{N}} \in \ell^1, \beta_{k+1} \rightarrow 0, \quad (2.1)$$

donde ℓ^1 viene de la Definición (1.1).

Paso 1: Dado $x^k \in H$, find $x^{k+1} \in H$ tal que:

$$g(x^{k+1}) + \theta \|x^{k+1} - x^k\|^2 \leq g(x^k) + \varepsilon_{k+1} \quad (2.2)$$

$$v^{k+1} \in \partial_{\varepsilon_{k+1}} g(x^{k+1}) \quad (2.3)$$

donde

$$\|v^{k+1}\| \leq \kappa \|x^{k+1} - x^k\| + \beta_{k+1}. \quad (2.4)$$

Paso 2: Si $0 \in \partial g(x^{k+1})$ or $\|x^k - x^{k+1}\| = 0$, entonces parar. Caso contrario, hacer $k \rightarrow k + 1$ y retornar al Paso 1.

El algoritmo abstracto inexacto introducido se convierte en el algoritmo propuesto por Attouch et al. [22] cuando H es el espacio euclídeo, $\varepsilon_k = 0$ y $\beta_k = 0$. En la siguiente sección presentaremos especializaciones de nuestro Algoritmo 1.

2.2. Algoritmos con propiedad Epsilon-Descenso

2.2.1. Método de gradiente inexacto con errores

Consideramos una función $g : H \rightarrow \mathbb{R}$ de clase C^1 con gradiente L -Lipschitz, i.e. ∇g es L -Lipschitz.

Algoritmo 2 Método de gradiente inexacto con errores

Paso 0: Tomar dos números reales θ, κ con $\theta > L$. Elegir $x^0 \in H$ y dos sucesiones reales no negativas $\{\epsilon_{k+1}\}$ y $\{\beta_{k+1}\}$ tal que:

$$\{\epsilon_{k+1}\}_{k \in \mathbb{N}}, \{\beta_{k+1}\}_{k \in \mathbb{N}} \in \ell^1. \quad (2.5)$$

Paso 1: Dado $x^k \in H$, encontrar $x^{k+1} \in H$ tal que:

$$\frac{\theta}{2} \|x^{k+1} - x^k\|^2 \leq \langle \nabla g(x^k), x^k - x^{k+1} \rangle + \epsilon_{k+1} \quad (2.6)$$

$$\|\nabla g(x^k)\| \leq \kappa \|x^{k+1} - x^k\| + \beta_{k+1}. \quad (2.7)$$

Paso 2: Si $\|x^k - x^{k+1}\| = 0$ o $\nabla g(x^{k+1}) = 0$, entonces parar. Caso contrario, hacer $k \rightarrow k + 1$ y retornar al Paso 1.

Observación 2.2.1 Cuando $\epsilon_{k+1} = 0$ y H es de dimensión finita, es Algoritmo 2 se reduce al algoritmo introducido en la sección 2 de Attouch et al. [22].

Observación 2.2.2 Sea a y b dos números reales positivos y $\{\gamma^k\}$ tal que $0 < a \leq \gamma^k \leq b < \frac{2}{L}$, donde L es la constante de Lipschitz de la gradiente de g . Consideremos una sucesión generada por el método del gradiente inexacto con un error.

$$x^{k+1} = x^k - \gamma^k \nabla g(x^k) + e^k \quad (2.8)$$

Obsérvese que en las aplicaciones e^k modelan errores de medida, ruidos, errores de entrada, aproximaciones, entre otros. Suponemos que

$$\sum_{k=1}^{+\infty} \|e^k\| < +\infty. \quad (2.9)$$

Afirmamos que el método del gradiente (2.8)-(2.9) es un caso particular del algoritmo 2. De hecho, de (2.8) tenemos

$$\|x^{k+1} - x^k\|^2 = \|- \gamma^k \nabla g(x^k) + e^k\|^2$$

$$\begin{aligned}
&= \langle -\gamma^k \nabla g(x^k) + e^k, x^{k+1} - x^k \rangle = \gamma^k \langle \nabla g(x^k), x^k - x^{k+1} \rangle + \langle e^k, x^{k+1} - x^k \rangle \\
&\leq \gamma^k \langle \nabla g(x^k), x^k - x^{k+1} \rangle + \|e^k\| \|x^{k+1} - x^k\|
\end{aligned} \tag{2.10}$$

Notamos que $ab \leq \frac{a^2+b^2}{2}$, para todo $a, b \in \mathbb{R}$ tenemos

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

Usando la anterior desigualdad en (2.10)

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

Esto nos da

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

Utilizando el hecho $0 < a \leq \gamma^k \leq b < \frac{2}{L}$ tenemos

$$\frac{1}{2b} \|x^{k+1} - x^k\|^2 \leq \langle \nabla g(x^k), x^k - x^{k+1} \rangle + \frac{\|e^k\|^2}{2a} \tag{2.11}$$

Tomando $\epsilon_{k+1} = \frac{\|e^k\|^2}{2a}$, la Condición (2.6) es satisfecha.

Ahora, mostramos que la Condición (2.7) es también cumplida. De (2.8)

$$\|\nabla g(x^k)\| = \frac{1}{\gamma^k} \|x^{k+1} - x^k - e^k\| \leq \frac{1}{a} \|x^{k+1} - x^k\| + \frac{1}{a} \|e^k\| \tag{2.12}$$

Tomando $\beta_{k+1} = \frac{1}{a} \|e^k\|$ tenemos la Condición (2.7).

Ahora, vamos a probar que el Algoritmo 2 cae en la forma del algoritmo abstracto generalizado de epsilon-descent, Algoritmo 1 .

La Condición (2.1) del Algoritmo 1 se obtiene arriba. Para demostrar que se cumple la

condición (2.2), aplicamos el Lema 1.1.13 en los puntos $x = x^k$ y $y = x^{k+1}$ y añadiendo el término $\frac{\theta}{2}\|x^{k+1} - x^k\|^2$ en ambos miembros de la desigualdad, entonces obtenemos:

$$\langle \nabla g(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 + \frac{\theta}{2}\|x^{k+1} - x^k\|^2 \geq g(x^{k+1}) - g(x^k) + \frac{\theta}{2}\|x^{k+1} - x^k\|^2.$$

Por (2.6):

$$g(x^{k+1}) + \frac{\theta}{2}\|x^{k+1} - x^k\|^2 \leq g(x^k) + \frac{L}{2}\|x^{k+1} - x^k\|^2 + \varepsilon_{k+1}$$

Entonces,

$$g(x^{k+1}) + \left(\frac{\theta - L}{2}\right)\|x^{k+1} - x^k\|^2 \leq g(x^k) + \varepsilon_{k+1}.$$

Ya que $\theta > L$ tenemos (2.2).

Por la Proposición 1.3.11 obtenemos:

$$\partial_{\varepsilon_{k+1}}g(x^{k+1}) = \{\nabla g(x^{k+1}) + \varepsilon_{k+1}B^*\}.$$

Entonces, para algún $z \in B$, tenemos que $v^{k+1} = \nabla g(x^{k+1}) + \varepsilon_{k+1}z \in \partial_{\varepsilon_{k+1}}g(x^{k+1})$, y ya que ∇g es L -Lipschitz, aplicando la desigualdad triangular obtenemos que:

$$\begin{aligned} \|v^{k+1}\| &= \|\nabla g(x^{k+1}) + \varepsilon_{k+1}z\| \leq \|\nabla g(x^{k+1})\| + \|\varepsilon_{k+1}z\| \\ &\leq \|\nabla g(x^{k+1}) - \nabla g(x^k)\| + \|\nabla g(x^k)\| + \varepsilon_{k+1}\|z\| \\ &\leq (L + \kappa)\|x^{k+1} - x^k\| + \varepsilon_{k+1} + \beta_{k+1}. \end{aligned}$$

Por lo tanto, se cumple la condición (2.4) del Algoritmo 1.

2.2.2. Método proximal inexacto

El método del punto proximal para encontrar un punto mínimo de una función semicontinua inferior adecuada $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$, partiendo de un punto x^0 , genera una sucesión $\{x^k\}$ dada por la siguiente iteración

$$x^{k+1} \in \arg \min \left\{ g(x) + \frac{\lambda_k}{2} \|x - x^k\|^2 : x \in H \right\}. \quad (2.13)$$

donde λ_k es un parámetro positivo. Gracias al Proposición 1.1.12 sabemos que para asegurar la existencia de cada x^k es g debe ser acotada inferiormente y semicontinua inferior débil.

Suponiendo la existencia de x^{k+1} tenemos de la condición de minimizador y de la condición de optimalidad que

$$\begin{aligned} g(x^{k+1}) &\leq g(x^k) - \frac{\lambda_k}{2} \|x^{k+1} - x^k\|^2 \\ 0 &\in \hat{\partial} \left(g(\cdot) + \frac{\lambda_k}{2} \|\cdot - x^k\|^2 \right) (x^{k+1}) \end{aligned}$$

Las dos condiciones anteriores se explotarán para definir un algoritmo de punto proximal inexacto.

Un tipo de algoritmo proximal inexacto

El siguiente algoritmo extiende el algoritmo introducido por Attouch et al. [22], Sección 4, de \mathbb{R}^n a H considerando el error escalar en la condición de descenso y el error relativo. Este algoritmo se puede aplicar a una función propia $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$, que es débilmente semicontinua inferior y acotada inferiormente.

Algoritmo 3 Método proximal inexacto

Paso 0: Elegir $x^0 \in H$, $0 \leq \omega < 1$, $0 < \underline{\tau} \leq \bar{\tau} < \infty$, $0 < \delta \leq 1$, y una sucesión real no negativa $\{\varepsilon_{k+1}\}$ tal que:

$$\sum_{k=0}^{\infty} \varepsilon_{k+1} < +\infty. \quad (2.14)$$

Paso 1: Tomar $\tau_k \in [\underline{\tau}, \bar{\tau}]$, dado $x^k \in H$, encontrar $x^{k+1}, v^{k+1} \in H$ tal que:

$$g(x^{k+1}) \leq g(x^k) - \frac{\delta}{2\tau_k} \|x^k - x^{k+1}\|^2 + \varepsilon_{k+1} \quad (2.15)$$

$$v^{k+1} \in \partial_{\varepsilon_{k+1}} g(x^{k+1}) \quad (2.16)$$

$$\|\tau_k v^{k+1} + x^{k+1} - x^k\|^2 \leq \omega(\|\tau_k v^{k+1}\|^2 + \|x^{k+1} - x^k\|^2) \quad (2.17)$$

Paso 2: Si $0 \in \partial g(x^{k+1})$ or $\|x^k - x^{k+1}\| = 0$, entonces parar. De lo contrario, haga lo siguiente $k \rightarrow k + 1$ y volver al paso 1.

El algoritmo del punto proximal (2.13) para $\lambda_k = \frac{\delta}{\tau_k}$ es un caso particular del algoritmo anterior. De hecho, dado $x^k \in H$, observamos que

- Es evidente que existe x^{k+1} solución del problema $\min\{g(x) + \frac{\delta}{2\tau_k} \|x^k - x\|^2 : x \in H\}$, por lo mencionado al inicio de esta sección y por la Proposición 1.1.12. Luego existirá un ε -mínimo del problema; entonces, para todo $x \in H$, x^{k+1} tiene:

$$g(x) + \frac{\delta}{2\tau_k} \|x^k - x\|^2 + \varepsilon_{k+1} \geq g(x^{k+1}) + \frac{\delta}{2\tau_k} \|x^k - x^{k+1}\|^2.$$

Tomando $x = x^k$ obtenemos:

$$g(x^k) - \frac{\delta}{2\tau_k} \|x^k - x^{k+1}\|^2 + \varepsilon_{k+1} \geq g(x^{k+1}).$$

- Por la proposición 1.3.5, $0 \in \partial(g(\cdot) + \frac{\delta}{2\tau_k} \|x^k - \cdot\|^2)(x^{k+1})$ y por Proposición 1.3.8, obtenemos las condiciones (2.16) y (2.17) del algoritmo anterior.

De (2.14), (2.15) y (2.16) obtenemos directamente las condiciones (2.1), (2.2) y (2.3) del Algoritmo 1. Para demostrar que se cumple la condición (2.4), necesitamos utilizar el siguiente lema, que es una extensión a los espacios de Hilbert del lema 4.1, página 113 de [22].

Lema 2.2.1 *Dado $x, y \in H$ y $\rho \in (0, 1]$ tal que $\rho(\|x\|^2 + \|y\|^2) \geq \|x+y\|^2$. Se cumple la siguiente desigualdad*

$$2\langle x, y \rangle \leq (\rho - 1)(\|x\|^2 + \|y\|^2).$$

Además, si también $\rho \in (0, 1)$, entonces:

$$\alpha^{-1}(1 + \sqrt{1 - \alpha^2})\|y\| \geq \|x\| \geq \alpha^{-1}(1 - \sqrt{1 - \alpha^2})\|y\| \quad (2.18)$$

donde $\alpha = 1 - \rho$.

Demostración:

Similar a la prueba del Lema 4.1 de Attouch et al. [22]. ■

Como $\omega < 1$, considerando $\rho = \omega$, la Desigualdad (2.17) con $x = \tau_k v^{k+1}$ y $y = x^{k+1} - x^k$ y usando el Lema 2.2.1, la desigualdad de la izquierda en (2.18), obtenemos para algún $\kappa > 0$:

$$\kappa \|x^{k+1} - x^k\| \geq \|\tau_k v^{k+1}\|,$$

entonces, ya que $\tau_k \in [\underline{\tau}, \bar{\tau}]$:

$$\|v^{k+1}\| \leq \frac{\kappa}{\tau_k} \|x^{k+1} - x^k\| \leq \frac{\kappa}{\underline{\tau}} \|x^{k+1} - x^k\|$$

y la Condición (2.4) se cumple.

Algoritmo proximal inexacto con cuasidistancias

Extendemos el algoritmo desarrollado en la Sección 4 por Papa Quiroz et al. [5] a los espacios de Hilbert. Primero consideramos algunas definiciones y proposiciones adicionales.

Definición 2.2.1 *Dado un conjunto V . Una cuasi distancia en V es un mapeo $\phi : V \times V \rightarrow \mathbb{R}_+$, tal que para cada $r, s, t \in V$,*

$$(i) \quad \phi(r, s) = \phi(s, r) = 0 \text{ siempre que } s = r,$$

$$(ii) \quad \phi(r, s) \leq \phi(r, t) + \phi(t, s).$$

Sea $\phi : V \times V \rightarrow \mathbb{R}_+$ sea una cuasidistancia. La cuasidistancia se llama no tan asimétrica si para cada $r, s \in V$ y algunos $\alpha, \beta > 0$, se mantiene:

$$\alpha \|r - s\| \leq \phi(r, s) \leq \beta \|r - s\| \quad (2.19)$$

Las principales propiedades de las cuasidistancias pueden encontrarse en Moreno et al. [44]. Aquí esbozamos algunas de ellas cuando la cuasidistancia satisface la Condición (2.19).

- Para todo $\bar{\omega} \in H$ las funciones $\phi(\bar{\omega}, \cdot)$ y $\phi(\cdot, \bar{\omega})$ son Lipschitz
- Para todo $\bar{\omega} \in H$ $\phi^2(\bar{\omega}, \cdot)$ y $\phi^2(\cdot, \bar{\omega})$ son funciones localmente Lipschitz y coercivas.

Para una función extendida, propia, débilmente semicontinua inferior, coerciva $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$, elegimos una cuasi-distancia $\phi : H \times H \rightarrow \mathbb{R}_+$ tal que satisfaga la Condición (2.19). Consideramos el siguiente algoritmo:

Algoritmo 4

Paso 0: Elegir ϕ una cuasi-distancia en H , $x^0 \in H$, $\rho \in \mathbb{R}_{++}$ y un par de sucesiones reales positivas $\{\mu_{k+1}\}$ y $\{\varepsilon_{k+1}\}$ tal que:

$$\{\varepsilon_{k+1}\}_{k \in \mathbb{N}} \in \ell^1. \quad (2.20)$$

Paso 1: Dado $x^k \in H$, encontrar $x^{k+1}, w^{k+1} \in H$ tal que:

$$g(x^{k+1}) + \frac{\mu_{k+1}}{2} \phi^2(x^k, x^{k+1}) \leq g(x^k) + \varepsilon_{k+1} \quad (2.21)$$

$$w^{k+1} \in \partial_{\varepsilon_{k+1}} f(x^{k+1}) \quad (2.22)$$

$$\|e^{k+1}\| \leq \rho \phi(x^k, x^{k+1}) \quad (2.23)$$

donde

$$e^{k+1} = w^{k+1} + \mu_{k+1} \phi(x^k, x^{k+1}) \psi^{k+1} \quad (2.24)$$

$$\psi^{k+1} \in \partial \phi(x^k, \cdot)(x^{k+1}).$$

Paso 2: Si $x^k = x^{k+1}$ o $0 \in \partial g(x^{k+1})$, entonces parar. De lo contrario, hacer $k \rightarrow k+1$ y volver al Paso 1.

Dado $x^k \in H$, tenemos que el algoritmo está bien definido (x^{k+1} y w^{k+1} existen satisfaciendo todas la condiciones del algoritmo) porque:

- (2.21) se cumple porque el problema $\min\{f(x) + \frac{\mu_{k+1}}{2} \phi^2(x^k, x) : x \in H\}$ tiene solución, esto es debido a que la función a minimizar es una función coerciva, débilmente semi-continua inferior, y aplicando la Proposición 1.1.12,
- Ahora, por la Proposición 1.3.5, $0 \in \hat{\partial}(f(\cdot) + \frac{\mu_{k+1}}{2} \phi^2(x^k, \cdot))(x^{k+1})$ y por la Proposición 1.3.8, (2.22)-(2.24) se cumplen cuando $e^{k+1} = 0$.

Bajo algunas condiciones extra verificaremos que el algoritmo cae dentro de nuestro algoritmo abstracto inexacto. De (2.24) y la desigualdad triangular tenemos:

$$\|w^{k+1}\| \leq \|e^{k+1} - \mu_{k+1} \phi(x^k, x^{k+1}) \psi^{k+1}\| \leq \|e^{k+1}\| + \|\mu_{k+1} \phi(x^k, x^{k+1}) \psi^{k+1}\|,$$

$$\leq \rho\phi(x^k, x^{k+1}) + \mu_{k+1} \phi(x^k, x^{k+1}) \|\psi^{k+1}\| = (\rho + \mu_{k+1} \|\psi^{k+1}\|) \phi(x^k, x^{k+1}).$$

Así obtenemos:

$$\|w^{k+1}\| \leq (\rho + \mu_{k+1} \|\psi^{k+1}\|) \phi(x^k, x^{k+1}).$$

Ahora, cada vez que $\mu_{k+1} \|\psi^{k+1}\| \leq M$, donde M es una constante, tenemos la desigualdad:

$$\|w^{k+1}\| \leq (\rho + M) \phi(x^k, x^{k+1}),$$

y desde que $\varepsilon_k \geq 0$, $\forall k \in \mathbb{N}$ obtenemos:

$$\|w^{k+1}\| \leq (\rho + M) \phi(x^k, x^{k+1}) + \varepsilon_{k+1}.$$

Si consideramos la cuasi-distancia como la distancia euclídea, tenemos que el Algoritmo 4 satisface la condición (2.4) de nuestro algoritmo abstracto inexacto, las condiciones (2.1), (2.2) y (2.3) se obtienen directamente de (2.20), (2.21) y (2.22).

2.2.3. Algoritmo inexacto de división hacia adelante y atrás con errores

Consideremos una función propia débilmente semicontinua inferior y acotada inferiormente $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ que puede expresarse en la forma

$$g = \phi + \varphi,$$

donde $\phi : H \rightarrow \mathbb{R}$ es una función no lisa y no convexa y $\varphi : H \rightarrow \mathbb{R}$ es una función C^1 tal que $\nabla\varphi$ es L -Lipschitz. Este tipo de problema estructurado se encuentra con frecuencia (véase por ejemplo [23] para problemas de sistemas de inclusiones monótonas acopladas en espacios de Hilbert y [34] para problemas de recuperación de señales).

Considere el siguiente algoritmo:

Algoritmo 5

Paso 0: Tomar $\theta, \kappa > 0$ con $\theta > L$. Establecer un punto inicial $x^0 \in \text{dom}(\phi)$ y una sucesión real no negativa $\{\varepsilon_{k+1}\}$ tal que:

$$\sum_{k=0}^{+\infty} \varepsilon_{k+1} < +\infty. \quad (2.25)$$

Paso 1: Dado $x^k \in H$, encontrar $x^{k+1}, v^{k+1} \in H$ tal que:

$$\phi(x^{k+1}) + \langle x^{k+1} - x^k, \nabla\phi(x^k) \rangle + \frac{\theta}{2} \|x^{k+1} - x^k\|^2 \leq \phi(x^k) + \varepsilon_{k+1} \quad (2.26)$$

$$v^{k+1} \in \partial_{\varepsilon_{k+1}}\phi(x^{k+1}) \quad (2.27)$$

$$\|v^{k+1} + \nabla\phi(x^k)\| \leq \kappa \|x^{k+1} - x^k\| \quad (2.28)$$

Paso 2: Si $\|x^k - x^{k+1}\| = 0$ or $0 \in \partial g(x^{k+1})$, entonces parar. De lo contrario, hacer $k \rightarrow k + 1$ y volver al Paso 1.

Observación 2.2.3 *El método de división hacia delante y hacia atrás*

$$x^{k+1} \in \text{prox}_{\lambda_k\phi}(x^k - \lambda_k \nabla\phi(x^k)).$$

donde $\text{prox}_{\lambda f}(x) := \arg \min\{f(y) + \frac{1}{2\lambda}\|y - x\|^2 : y \in H\}$ y $\lambda_k > 0$, es un ejemplo del Algoritmo 5 considerando en cada iteración $\varepsilon_{k+1} = 0$. De hecho, esa afirmación fue demostrada por Attouch et al. [22], subsección 5.1, bajo la introducción de dos números positivos $\underline{\lambda}, \bar{\lambda}$ tal que: $0 < \underline{\lambda} < \lambda_k < \bar{\lambda} < \frac{1}{L}$.

Ahora demostramos que el Algoritmo 5 satisface las condiciones de nuestro algoritmo abstracto inexacto (Algoritmo 1). Aplicamos el Lema 1.1.13 a la función ϕ con $y = x^{k+1}$ y $x = x^k$:

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

Añadiendo en ambos miembros $\phi(x^{k+1}) + \frac{\theta}{2}\|x^{k+1} - x^k\|^2$ y por (2.26):

$$\begin{aligned} & \phi(x^{k+1}) + \phi(x^{k+1}) + \frac{\theta}{2}\|x^{k+1} - x^k\|^2 \\ & \leq \phi(x^k) + \langle x^{k+1} - x^k, \nabla\phi(x^k) \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 + \phi(x^{k+1}) + \frac{\theta}{2}\|x^{k+1} - x^k\|^2 \\ & \leq \phi(x^k) + \varepsilon_{k+1} + \phi(x^k) + \frac{L}{2}\|x^{k+1} - x^k\|^2. \end{aligned}$$

Después de algunas operaciones algebraicas, tenemos:

$$\phi(x^{k+1}) + \phi(x^{k+1}) + \left(\frac{\theta - L}{2}\right)\|x^{k+1} - x^k\|^2 \leq \phi(x^k) + \phi(x^k) + \varepsilon_{k+1}.$$

Dado que $g = \phi + \phi$, se cumple:

$$g(x^{k+1}) + \left(\frac{\theta - L}{2}\right)\|x^{k+1} - x^k\|^2 \leq g(x^k) + \varepsilon_{k+1}.$$

Es decir, se cumple la condición (2.2) del Algoritmo 1.

Ahora, definimos $w^{k+1} = v^{k+1} + \nabla\phi(x^{k+1})$. Lo demostraremos:

$$w^{k+1} \in \partial_{\varepsilon_{k+1}}g(x^{k+1}). \quad (2.29)$$

De hecho, a partir del Algoritmo 5 sabemos que $v^{k+1} \in \partial_{\varepsilon_{k+1}}\phi(x^{k+1})$. Por lo tanto, por definición 1.3.6:

$$\exists x^l \longrightarrow x^{k+1}, \phi(x^l) \longrightarrow \phi(x^{k+1}), v^l \in \hat{\partial}_{\varepsilon_{k+1}}\phi(x^l) \text{ y } v^l \rightharpoonup v^{k+1}.$$

Dado que φ es continua, tenemos:

$$\exists x^l \longrightarrow x^{k+1} : g(x^l) = \phi(x^l) + \varphi(x^l) \longrightarrow \phi(x^{k+1}) + \varphi(x^{k+1}) = g(x^{k+1}). \quad (2.30)$$

Por la Proposición 1.3.9, ítem (3), se deduce que

$$\hat{\partial}_{\varepsilon_{k+1}}\phi(x^l) + \hat{\partial}_0\varphi(x^l) \subset \hat{\partial}_{\varepsilon_{k+1}}(\phi + \varphi)(x^l) = \hat{\partial}_{\varepsilon_{k+1}}g(x^l). \quad (2.31)$$

Entonces, por la Proposición 1.3.1, obtenemos $\nabla\varphi(x^l) \in \hat{\partial}_0\varphi(x^l) = \hat{\partial}\varphi(x^l)$ además $v^l \in \hat{\partial}_{\varepsilon_{k+1}}\phi(x^l)$, entonces de (2.31) obtenemos:

$$v^l + \nabla\varphi(x^l) \in \hat{\partial}_{\varepsilon_{k+1}}g(x^l). \quad (2.32)$$

Ya que $\nabla\varphi$ es continua y $v^l \rightarrow v^{k+1}$, por Proposición 1.1.9:

$$v^l + \nabla\varphi(x^l) \rightarrow v^{k+1} + \nabla\varphi(x^{k+1}) = w^{k+1}. \quad (2.33)$$

De (2.30), (2.32) y (2.33) concluimos que:

$$w^{k+1} \in \partial_{\varepsilon_{k+1}}g(x^{k+1}).$$

Por la Desigualdad (2.28) del Algoritmo 5, y la desigualdad triangular obtenemos la Condición (2.4) del Algoritmo 1, como sigue:

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

$$= (\kappa + L)\|x^{k+1} - x^k\|.$$

Entonces, ya que $\varepsilon_{k+1} \geq 0$ tenemos:

$$\|w^{k+1}\| \leq (\kappa + L)\|x^{k+1} - x^k\| + \varepsilon_{k+1}.$$

2.2.4. Un método de Gauss-Seidel regularizado inexacto

El siguiente algoritmo extiende el algoritmo introducido en Attouch et al. (2013) Sección 6 [22] a los espacios de Hilbert y considera errores escalares.

Antes de introducir el algoritmo es necesario revisar el operador proximal en espacios de Hilbert [19] cuando se utiliza nuevas métricas. Para ello consideraremos el espacio de operadores acotados, uniformemente elípticos y autoadjuntos sobre H , denotado por $S_{++}(H)$. Para cualquier $A \in S_{++}(H)$, el operador proximal para una función $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ se define:

$$\text{prox}_f^A(x) = \arg \min_{y \in H} \left\{ f(y) + \frac{1}{2} \langle A(y-x), y-x \rangle \right\}$$

Ahora, seguimos con el algoritmo y para ello elegimos un entero $p > 0$ y consideramos H_1, \dots, H_p espacios de Hilbert con $\langle \cdot, \cdot \rangle_{H_i}$ como producto interior y $\|\cdot\|_{H_i}$ como la norma inducida. Además $H = \prod_{i=1}^p H_i$ con $\langle \cdot, \cdot \rangle = \sum_{i=1}^p \langle \cdot, \cdot \rangle_{H_i}$ y norma $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

Consideremos el problema:

$$\min_{x_i \in H_i} f(x_1, \dots, x_p) = h(x_1, \dots, x_p) + \sum_{i=1}^p g_i(x_i),$$

donde $h : H \rightarrow \mathbb{R}$ es de clase C^1 cuya gradiente es localmente Lipschitz continua y cada $g_i : H_i \rightarrow \mathbb{R} \cup \{+\infty\}$ es una función propia semicontinua inferior débil. Además es necesario que f sea acotada inferiormente.

Para cada $i \in \{1, \dots, p\}$, vamos a considerar sucesiones acotadas de matrices $\{B_i^k\}_{k \in \mathbb{N}} \in S_{++}(H)$, donde $\underline{\lambda}$ y $\bar{\lambda}$ son cotas inferior y superior de los valores espectrales de las matrices $\{B_i^k : k \in \mathbb{N}, i \in \{1, \dots, p\}\}$ respectivamente.

Ahora, a partir de $x^k \in H$ construimos x^{k+1} componente a componente:

- $x_1^{k+1} \in \arg \min_{y \in H_1} \left\{ f(y, x_2^k, \dots, x_p^k) + \frac{1}{2} \langle A_1^k (y - x_1^k), y - x_1^k \rangle_{H_1} \right\}$
- Para $i = 2, \dots, p-1$:

$$x_i^{k+1} \in \arg \min_{y \in H_i} \left\{ f(x_1^{k+1}, \dots, x_{i-1}^{k+1}, y, x_{i+1}^k, \dots, x_p^{k+1}) + \frac{1}{2} \langle A_i^k (y - x_i^k), y - x_i^k \rangle_{H_i} \right\}$$

- $x_p^{k+1} \in \arg \min_{y \in H_p} \left\{ f(x_1^{k+1}, \dots, x_{p-1}^{k+1}, y) + \frac{1}{2} \langle A_p^k (y - x_p^k), y - x_p^k \rangle_{H_p} \right\}$

hacemos $x^{k+1} = (x_1^{k+1}, \dots, x_p^{k+1})$.

Siguiendo el razonamiento de la Proposición 1.1.12, como f es semicontinua inferior débil y acotada inferiormente x_i^{k+1} existe para $i = 1, \dots, p$ y por ende se garantiza la existencia de x^{k+1} .

Luego, aplicando la definición del operador proximal definida al inicio de esta sección y siguiendo el razonamiento que se utilizó cuando estudiamos los algoritmos proximales, se construye el algoritmo de Gauss Seidel con errores de la siguiente manera:

Algoritmo 6 Algoritmo inexacto de Gauss-Seidel con errores

Paso 1 Tomar $0 < \underline{\lambda} < \bar{\lambda} < \infty$. Para cada $i \in \{1, \dots, p\}$, tomar una sucesión acotada de matrices $\{A_i^k\}_{k \in \mathbb{N}} \in S_{++}(H)$ tal que los valores espectrales 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 $x^0 = (x_1^0, \dots, x_p^0) \in H$. Tomar dos sucesión de números reales positivos $\{\varepsilon_{k+1}\}$ y $\{\beta_{k+1}\}$ tal que:

$$\varepsilon_{k+1}, \beta_{k+1} \in \ell^1. \quad (2.34)$$

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

$$\begin{aligned} g_i(x_i^{k+1}) + h(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, \dots, x_p^k) + \frac{1}{2} \langle A_i^k(x_i^{k+1} - x_i^k), x_i^{k+1} - x_i^k \rangle_{H_i} \\ \leq g_i(x_i^k) + h(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_p^k) + \frac{\varepsilon_{k+1}}{p} \end{aligned} \quad (2.35)$$

$$v_i^{k+1} \in \partial_{\varepsilon_{k+1}} g_i(x_i^{k+1}) \quad (2.36)$$

$$\|v_i^{k+1} + \nabla_{x_i} h(x_1^{k+1}, \dots, x_i^{k+1}, x_{i+1}^k, \dots, x_p^k)\|_{H_i} \leq b_i \|x_i^{k+1} - x_i^k\|_{H_i} + \frac{\beta_{k+1}}{p} \quad (2.37)$$

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 $w^{k+1} = (v_i^{k+1} + \nabla_{x_i} h(x_1^{k+1}, \dots, x_p^{k+1}))_{i=1, \dots, p} \in H$ y afirmamos lo siguiente:

$$w^{k+1} \in \partial_{\varepsilon_{k+1}} f(x^{k+1}) \quad (2.38)$$

En efecto, de (2.36) sabemos que para cada i existe $v_i^{k+1} \in \partial_{\varepsilon_{k+1}} g_i(x_i^{k+1})$; entonces, usando las reglas de diferenciación para funciones separadas, tenemos:

$$v^{k+1} = (v_1^{k+1}, \dots, v_p^{k+1}) \in \partial_{\varepsilon_{k+1}} \sum_{i=1}^p g_i(x_i^{k+1}) \quad (2.39)$$

y como h es de clase C^1 , gracias la demostración de la afirmación (2.29) en la sección anterior, obtenemos que:

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

y así demostramos (2.38).

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

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

Por lo tanto, para algún $M > 0$ se cumple:

$$\|w^{k+1}\| \leq M \|x^{k+1} - x^k\| + \beta_{k+1} \quad (2.40)$$

Hasta ahora, se cumplen las condiciones (2.1), (2.3) y (2.4) del algoritmo 1. Vamos a demostrar que también se cumple la condición (2.2). Para ello usamos la desigualdad $\underline{\lambda} \|u\|_{H_i}^2 \leq \langle A_i^k u, u \rangle_{H_i}$ (explicada en [19]) para los enteros i y k en (2.36) y obtenemos:

$$\begin{aligned}
& g_i(x_i^{k+1}) + h(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, \dots, x_p^k) + \frac{\lambda}{2} \|x_i^{k+1} - x_i^k\|_{H_i}^2 \\
& \leq g_i(x_i^{k+1}) + h(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, \dots, x_p^k) + \frac{1}{2} \langle A_i^k (x_i^{k+1} - x_i^k), x_i^{k+1} - x_i^k \rangle_{H_i} \\
& \leq g_i(x_i^k) + h(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_p^k) + \frac{\epsilon_{k+1}}{p}
\end{aligned}$$

Considerando solo los extremos:

$$g_i(x_i^{k+1}) + h(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, \dots, x_p^k) + \frac{\lambda}{2} \|x_i^{k+1} - x_i^k\|_{H_i}^2$$

$$\leq g_i(x_i^k) + h(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_p^k) + \frac{\epsilon_{k+1}}{p} \quad (2.41)$$

Sumando las desigualdades de (2.41) desde $i = 1$ hasta $i = p$, obtenemos:

$$\sum_{i=1}^p g_i(x_i^{k+1}) + h(x^{k+1}) + \frac{\lambda}{2} \sum_{i=1}^p \|x_i^{k+1} - x_i^k\|_{H_i}^2 \leq \sum_{i=1}^p g_i(x_i^k) + h(x^k) + \epsilon_{k+1}$$

$$f(x^{k+1}) + \frac{\lambda}{2} \sum_{i=1}^p \langle x_i^{k+1} - x_i^k, x_i^{k+1} - x_i^k \rangle_{H_i} \leq f(x^k) + \epsilon_{k+1}$$

$$f(x^{k+1}) + \frac{\lambda}{2} \langle x^{k+1} - x^k, x^{k+1} - x^k \rangle_{H_i} \leq f(x^k) + \epsilon_{k+1}$$

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

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

Observación 2.2.4 *En el artículo introducido por Frankel et al. [19], Sección 4, desarrollaron un algoritmo del tipo división hacia adelante y hacia atrás y Gauss-Seidel, que gracias a su naturaleza, caen dentro del algoritmo abstracto inexacto introducido en este trabajo.*

Antes de continuar con el algoritmo anterior, debemos introducir la métrica dada del operador proximal presentada en la Sección 2.3 de [19].

Consideramos el espacio de operadores acotados, uniformemente elípticos y autoadjuntos en H denotado por $S_{++}(H)$. Cada $A \in S_{++}(H)$ induce una métrica sobre H mediante la norma $\|x\|_A = \sqrt{\langle x, x \rangle_A}$ donde el producto interno es definido por $\langle x, y \rangle_A = \langle A(x), y \rangle$. Además, denotamos $\alpha(A)$ como el ínfimo de los valores espectrales de A , tal que $\|x^2\|_A \geq \alpha(A)\|x\|^2 \forall x \in H$. Dado $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ el operador proximal de g en la métrica inducida por A es el mapeo $\text{prox}_g^A : H \rightarrow H$, definido como:

$$\text{prox}_g^A(x) = \operatorname{argmin}_{y \in H} \left\{ g(y) + \frac{1}{2} \|y - x\|_A^2 \right\}$$

Los espacios H y H_i son los mismos que consideramos al inicio de esta subsección. El problema a resolver es:

$$\min_{x_i \in H_i} f(x_1, \dots, x_p) = h(x_1, \dots, x_p) + \sum_{i=1}^p g_i(x_i),$$

donde $h : H \rightarrow \mathbb{R}$ es continuamente diferenciable y cada $g_i : H_i \rightarrow \mathbb{R}$ es una función semicontinua inferior. Tomar $L \geq 0$ tal que las funciones $x \in H_i \mapsto h(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_p)$ tienen gradiente L -Lipschitz continuo. Para la sucesión generada para el siguiente algoritmo, denotamos $X^k = (x_1^k, \dots, x_p^k)$ y $X_i^k = (x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_p^k)$. El algoritmo es presentado de la siguiente manera.

Algoritmo 7

Paso 1: Tomar $\alpha > L$, un entero $p > 0$, y para cada $i \in \{1, \dots, p\}$ elegir una sucesión $\{A_{i,k}\} \in S_{++}(H_i)$, donde α_k y β_k son el ínfimo y máximo de los valores espectrales de $\{A_{i,k}\}_{i \in \{1, \dots, p\}}$ tal que, $\alpha_k > \alpha$, $\frac{1}{\beta_k} \notin \ell^1$ y $\sup_{k \in \mathbb{N}} \frac{\beta_k}{\alpha_{k+1}} < +\infty$. Tomar $X^0 \in H$. Y elegir dos sucesiones $\{S^k = (s_1^k, \dots, s_p^k)\}, \{R^k = (r_1^k, \dots, r_p^k)\}$ tal que existe $\eta \geq 0, \mu \in]0, 1]$ con $\frac{\eta+1}{\mu} < \frac{\alpha}{L}$:

- $\|S_i^k\| \leq \frac{\eta}{2} \|y_i^{k+1} - y_i^k\|$, donde $S_i^k = (s_1^{k+1}, \dots, s_{i-1}^{k+1}, s_i^k, \dots, s_p^k)$
- $\|r_i^k\| \leq \frac{\eta}{2} \|y_i^{k+1} - y_i^k\| + \delta^k$, con $\delta_k \geq 0$ y $\delta_k \in \ell^1$,
- $\langle r_i^k + s_i^k, y_i^{k+1} - y_i^k \rangle_{A_{i,k}} \leq \frac{1-\mu}{2} \|y_i^{k+1} - y_i^k\|_{A_{i,k}}^2$.

Paso 2: Dado X^k , para cada $i = 1, \dots, p$, hacer:

$$y_i^{k+1} \in \text{prox}_{g_i}^{A_{i,k}} \left(x_i^k - A_{i,k}^{-1} \nabla_{ih}(X_i^k) + r_i^k \right)$$

$$x_i^{k+1} = y_i^{k+1} + s_i^{k+1}$$

Paso 3: Si $\|X^k - X^{k+1}\| = 0$ o $0 \in \partial f(X^{k+1})$, parar. Caso contrario, hacer $k \rightarrow k+1$ y regresar al Paso 2.

Capítulo III

Convergencia del algoritmo abstracto de épsilon-descenso generalizado

En este apartado analizaremos la convergencia de la sucesión generada por el Algoritmo 1, cuando la función en estudio es propia y semicontinua inferior.

Lema 3.0.1 (Lema 2 página 44 [41]) Sea $\{\mu_k\}, \{\chi_k\}, \{\alpha_k\} \subset \mathbb{R}_+$, tal que la sucesión satisfaga la desigualdad $\mu_{k+1} \leq (1 + \chi_k)\mu_k + \alpha_k$, con $\sum_{k=1}^{+\infty} \alpha_k < +\infty$ y $\sum_{k=1}^{+\infty} \chi_k < +\infty$. Entonces, $\{\mu_k\}$ es convergente.

Proposición 3.0.2 Dada una función propia semicontinua inferior $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$. La sucesión $\{x^k\}$ generada por algún algoritmo que satisface las condiciones del algoritmo abstracto (Algoritmo 1) satisface:

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

(ii) $\|x^{k+1} - x^k\| \rightarrow 0$.

(iii) Si $\{x^{k_l}\}$ es convergente a x^* entonces $\{x^{k_l+1}\}$ converge al mismo punto x^* .

Demostración:

(i) Por (2.2) obtenemos:

$$g(x^{k+1}) \leq g(x^k) + \varepsilon_{k+1}, \quad (3.1)$$

Denotando $\omega(x^k) = g(x^k) - \inf g(x)$ por (3.1) tenemos:

$$\omega(x^{k+1}) \leq \omega(x^k) + \varepsilon_{k+1}. \quad (3.2)$$

Como $\omega(x^k) \geq 0$, podemos aplicar el Lema 3.0.1, (2.1) y (3.2) para obtener que $\{\omega(x^k)\}$ converge y así obtenemos (i).

(ii) De (2.2) obtenemos:

$$\|x^{k+1} - x^k\| \leq \frac{1}{\theta}(g(x^k) - g(x^{k+1})) + \varepsilon_{k+1}. \quad (3.3)$$

Como $\{g(x^k)\}$ converge por el ítem (i) y $\{\varepsilon_{k+1}\}$ converge a 0 por (2.1), obtenemos:

$$\frac{1}{\theta}(g(x^k) - g(x^{k+1})) + \varepsilon_{k+1} \rightarrow 0.$$

Entonces, por (3.3) $\|x^{k+1} - x^k\|$ converge a cero.

(iii) Supongamos que $\{x^{k_l}\}$ converge a x^* , entonces usando la propiedad de desigualdad triangular obtenemos:

$$\|x^{k_l+1} - x^*\| \leq \|x^{k_l+1} - x^{k_l}\| + \|x^{k_l} - x^*\|.$$

Since $x^{k_l} \rightarrow x^*$ y de (ii) obtenemos $\|x^{k_l+1} - x^*\| \rightarrow 0$. Por tanto, $\{x^{k_l+1}\}$ converge a x^* .

■

Definimos el conjunto:

$$U = \left\{ x \in H : g(x) \leq \lim_{k \rightarrow +\infty} g(x^k) \right\}.$$

Proposición 3.0.3 *Dada una función propia semicontinua inferior $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$. Cada punto de acumulación de la sucesión $\{x^k\}$ generada por el Algoritmo 1, si existe, pertenece a U .*

Demostración:

Suponemos que existe $\{x^{k_l}\} \subset \{x^k\}$ que converge a \bar{x} . Por la semicontinuidad inferior de g y desde que $\{g(x^k)\}$ converge por la Proposición 3.0.2, ítem (i), obtenemos:

$$g(\bar{x}) \leq \lim_{l \rightarrow +\infty} g(x^{k_l}) = \lim_{k \rightarrow +\infty} g(x^k).$$

Entonces $\bar{x} \in U$. ■

Teorema 3.0.4 *Dado la función propia y semicontinua inferior $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ y además g es continua en su dominio. Cada punto de acumulación de la sucesión $\{x^k\}$ generada por el Algoritmo 1, si existe, es un punto crítico límite generalizado de g .*

Demostración:

Supongamos que existe $\{x^{k_l}\} \subset \{x^k\}$ que converge a \bar{x} . Por Proposición 3.0.2, (iii), $x^{k_l+1} \rightarrow \bar{x}$ y además $g(x^{k_l+1}) \rightarrow g(\bar{x})$ ya que g es continua en su dominio. De (2.3) y (2.4) tenemos:

$$w^{k_l+1} \in \partial_{\varepsilon_{k_l+1}} g(x^{k_l+1}) \text{ and } \|w^{k_l+1}\| \leq \kappa \|x^{k_l+1} - x^{k_l}\| + \beta_{k_l+1}.$$

Por Proposición 3.0.2, (ii), sabemos que $\|x^{k_l+1} - x^{k_l}\| \rightarrow 0$, y gracias a (2.1) tenemos que $\beta_{k+1} \rightarrow 0$, luego

$$\|w^{k_l+1}\| \rightarrow 0.$$

Así, tenemos que existe $x^{k_l+1} \rightarrow \bar{x}$ con $g(x^{k_l+1}) \rightarrow g(\bar{x})$, y $w^{k_l+1} \in \partial_{\varepsilon_{k_l+1}} g(x^{k_l+1})$ con $w^{k_l+1} \rightarrow 0$. Entonces, $0 \in \bar{\partial}g(\bar{x})$. ■

En las siguientes secciones estudiamos las propiedades de convergencia del algoritmo para dos ε -subdiferenciales: ε -subdiferencial de Clarke y ε -subdiferencial convexa. La motivación para hacer este estudio es encontrar mejores propiedades que en el caso general. La última sección es para el análisis de convergencia bajo el supuesto de la Desigualdad de ε -Kurdyka Lojasiewicz.

3.1. Algoritmo abstracto de ε -descenso generalizado para funciones localmente lipschitz

Utilizaremos la subdiferencial ε -Clarke, introducida en la Sección 5 de Papa Quiroz et. al. [5].

Definición 3.1.1 *El subdiferencial ε -Clarke de g en $x \in \text{dom}(g)$, denotado por $\partial_\varepsilon^\circ g(x)$, es el siguiente conjunto.*

$$\partial_\varepsilon^\circ g(x) = \{v \in H : g^\circ(x, z) \geq \langle v, u \rangle - \varepsilon, \forall u \in H\}, \quad (3.4)$$

donde

$$g^\circ(x, u) = \limsup_{t \downarrow 0, y \rightarrow x} \frac{g(y + tu) - g(y)}{t}.$$

Presentamos una variante de nuestro Algoritmo 1 cuando la función $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ es

localmente lipschitz.

Algoritmo 8 Algoritmo abstracto de ε -descenso para funciones localmente lipschitz

Paso 0: Elegir $x^0 \in H$, $\theta, \kappa \in \mathbb{R}$ constantes positivas y dos sucesiones no negativas $\{\varepsilon_{k+1}\}, \{\beta_{k+1}\} \subset \mathbb{R}$ tal que:

$$\{\varepsilon_{k+1}\}_{k \in \mathbb{N}} \in \ell^1, \beta_{k+1} \rightarrow 0 \quad (3.5)$$

Paso 1: Dado x^k encontrar x^{k+1} que cumpla:

$$g(x^{k+1}) + \theta \|x^{k+1} - x^k\|^2 \leq g(x^k) + \varepsilon_{k+1} \quad (3.6)$$

$$v^{k+1} \in \partial_{\varepsilon_{k+1}}^\circ g(x^{k+1}) \quad (3.7)$$

donde

$$\|v^{k+1}\| \leq \kappa \|x^{k+1} - x^k\| + \beta_{k+1} \quad (3.8)$$

Paso 2: If $\|x^k - x^{k+1}\| = 0$ o $0 \in \partial^\circ g(x^{k+1})$, parar. Caso contrario, hacer $k \rightarrow k + 1$ y retornar al Paso 1.

Proposición 3.1.1 Dado la función propia, semicontinua inferior y localmente lipschitziana $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ y $\{x^k\}$ una sucesión generada por el Algoritmo 8. Cada punto de acumulación de $\{x^k\}$, si existe, es un punto crítico de Clarke de g .

Demostración:

Asumimos que existe $\bar{x} \in H$ un punto de acumulación de $\{x^k\}$, entonces existe $\{x^{k_j}\}$ una subsucesión de $\{x^k\}$ que converge a \bar{x} . Por Proposición 3.0.2, (iii), $x^{k_j+1} \rightarrow \bar{x}$. De (3.7), $v^{k_j+1} \in \partial_{\varepsilon_{k_j+1}}^\circ g(x^{k_j+1})$ y así

$$g^\circ(x^{k_j+1}, z) \geq \langle v^{k_j+1}, z \rangle - \varepsilon_{k_j+1}, \forall z \in H. \quad (3.9)$$

Por la desigualdad de Cauchy-Schwarz y (3.8):

$$\langle v^{k_j+1}, z \rangle \leq \|v^{k_j+1}\| \|z\| \leq \kappa \|x^{k_j+1} - x^{k_j}\| \|z\| + \beta_{k_j+1} \|z\|.$$

Análogo al razonamiento utilizado en la demostración del teorema. 3.0.4, se sigue que $\|v^{k_j+1}\| \|z\| \rightarrow 0, \forall z \in H$. Por lo tanto, para todo $z \in H, \langle v^{k+1}, z \rangle \rightarrow 0$.

Finalmente, tomando \limsup en ambos miembros de la desigualdad (3.9) y por la semicontinuidad superiormente de $g^\circ(\cdot, \cdot)$, obtenemos que: $g^\circ(\bar{x}, z) \geq \limsup_{j \rightarrow +\infty} g^\circ(x^{k_j+1}, z) \geq 0$. Esto implica que $0 \in \partial^\circ g(\bar{x})$.

■

3.2. Algoritmo abstracto de ϵ -descenso generalizado para funciones convexas

Presentamos una variante de nuestro algoritmo abstracto inexacto para el caso convexo.

Algoritmo 9 Algoritmo abstracto de ϵ -descenso generalizado para funciones convexas

Paso 0: Elegir $x^0 \in H, \theta, \kappa \in \mathbb{R}$ constantes positivas y dos sucesiones reales no negativas $\{\epsilon_{k+1}\}, \{\beta_{k+1}\} \subset \mathbb{R}$ such that:

$$\{\epsilon_{k+1}\}_{k \in \mathbb{N}} \in \ell^1, \beta_{k+1} \rightarrow 0 \quad (3.10)$$

Paso 1: Dado x^k calcular x^{k+1} tal que:

$$g(x^{k+1}) \leq g(x^k) - \theta \|x^{k+1} - x^k\|^2 + \epsilon_{k+1} \quad (3.11)$$

$$v^{k+1} \in \partial_{\epsilon_{k+1}}^F g(x^{k+1}) \quad (3.12)$$

donde

$$\|v^{k+1}\| \leq \kappa \|x^{k+1} - x^k\| + \beta_{k+1} \quad (3.13)$$

Paso 2: Si $0 \in \partial^F g(x^{k+1})$ o $\|x^k - x^{k+1}\| = 0$, parar. Caso contrario, hacer $k \rightarrow k+1$ y regresar al Paso 1.

Proposición 3.2.1 Sea $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia, semicontinua inferior y convexa. Sea $\{x^k\}$ una sucesión generada por el Algoritmo 1. Si $\{x^k\}$ es acotada, entonces cada

punto de acumulación débil de $\{x^k\}$ es un mínimo global de g .

Demostración:

Como $\{x^k\}$ es acotada y de la Proposición 1.1.11 existe un punto de acumulación débil, el cual denotaremos por \bar{x} , existe una subsucesión $\{x^{k_j}\}$ que converge debilmente a \bar{x} . Como g es semicontinua inferior débil pues g es convexa y semicontinua inferior (revisar la Proposición 1.1.10), y ya que $\{g(x^k)\}$ converge por la Proposición 3.0.2, (i), obtenemos que:

$$g(\bar{x}) \leq \lim_{k \rightarrow +\infty} g(x^{k_j}) = \lim_{k \rightarrow +\infty} g(x^k). \quad (3.14)$$

Ahora, dado que $v^{k+1} \in \partial_{\varepsilon_{k+1}}^F g(x^{k+1})$, por la Definición 1.3.4, para todo $z \in H$ la siguiente desigualdad es satisfecha

$$g(z) \geq g(x^{k+1}) + \langle v^{k+1}, z - x^{k+1} \rangle - \varepsilon_{k+1}. \quad (3.15)$$

Gracias a las Condiciones (3.10), (3.13), a la Proposición 3.0.2, (ii), tenemos que $\|v^{k+1}\| \rightarrow 0$. Además, como $\{x^k\}$ es acotada, tenemos que $\langle v^{k+1}, z - x^{k+1} \rangle \rightarrow 0$. Usando el hecho (3.14) en (3.15), y añadiendo el hecho $\varepsilon_{k+1} \rightarrow 0$, obtenemos que

$$g(\bar{x}) \leq \lim_{k \rightarrow \infty} g(x^k) \leq g(z), \quad \forall z \in H. \quad (3.16)$$

Por tanto, $\bar{x} \in H$ es un punto mínimo de g .



Observación 3.2.1 Una condición suficiente para obtener que la sucesión $\{x^k\}$ sea acotada y obtener la existencia del punto de acumulación débil de $\{x^k\}$ es que la función g sea semicon-

tinua inferior y coerciva. En efecto, de (3.1), tenemos:

$$g(x^k) < \alpha_0 := g(x^0) + \sum_{i=1}^{+\infty} \varepsilon_i.$$

Así, $x^k \in L_g(\alpha_0)$ y como g es semicontinua inferior y coerciva, entonces $\{x^k\}$ es acotada. Ahora, de la Proposición 1.1.11 existe $\{x^{k_j}\}$ una subsucesión de $\{x^k\}$ que converge debilmente a algún $\bar{x} \in H$ i.e. \bar{x} es un punto de acumulación débil.

3.3. Análisis de convergencia para funciones con la propiedad de ε -Kurdyka Lojasiewicz

Vamos a considerar la nueva versión de la propiedad de Kurdyka Lojasiewicz donde se incluye el ε subdiferencial límite de Fréchet. (Definición 1.4.4). Consideraremos el siguiente algoritmo:

Algoritmo 10 Algoritmo abstracto de ε -descenso generalizado para funciones KL

Paso 0: Elegir $x^0 \in \mathbb{R}^n$, dos números reales positivos θ, κ tal que $3\sqrt{\theta} - \kappa > 0$ una sucesión monótona real no negativa $\{\varepsilon_{k+1}\}$ y una sucesión $\{\beta_{k+1}\} \subset \mathbb{R}_+$ tal que:

$$\sum_{k=0}^{\infty} \sqrt{\varepsilon_{k+1}} < +\infty \quad \sum_{k=0}^{\infty} \beta_k < +\infty \quad (3.17)$$

Paso 1: Dado $x^k \in \mathbb{R}^n$, encontrar $x^{k+1} \in \mathbb{R}^n$ tal que:

$$g(x^{k+1}) + \theta \|x^{k+1} - x^k\|^2 \leq g(x^k) + \varepsilon_{k+1} \quad (3.18)$$

tal que:

$$x^{k+1} \in \{x \in \mathbb{R}^n : g(x) < g(x^k)\}$$

$$v^{k+1} \in \partial_{\varepsilon_{k+1}} g(x^{k+1}) \quad (3.19)$$

donde

$$\|v^{k+1}\| \leq \kappa \|x^{k+1} - x^k\| + \beta_{k+1}. \quad (3.20)$$

Step 2: Si $0 \in \partial g(x^{k+1})$ o $\|x^k - x^{k+1}\| = 0$, parar. Caso contrario, hacer $k \rightarrow k + 1$ y retornar al Paso 1.

Lema 3.3.1 Sea $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia, semicontinua inferior que cumple con la desigualdad de ε -Kurdyka Lojasiewicz en x^* . Sea $\{x^k\}$ una sucesión originada por el Algoritmo 10. Y tomamos los objetos de la Definición 1.4.4. Sea $\delta, \rho > 0$ tal que $B(x^*, \delta) \subset U$ con $\rho \in (0, \delta)$ luego:

Si $\sqrt{\frac{\eta+\gamma}{\theta}} \leq \delta - \rho$ y $f(x^k) \geq f(x^*) \forall k \in \mathbb{N}$ entonces

$$\forall k \in \mathbb{N}, x^k \in B(x^*, \rho) \implies x^{k+1} \in B(x^*, \delta) \text{ con } g(x^{k+1}) \geq g(x^*)$$

Demostración:

Sea $x^k \in B(x^*, \rho)$ por la Condición 3.18 tenemos:

$$\|x^{k+1} - x^k\| \leq \sqrt{\frac{f(x^k) - f(x^{k+1}) + \varepsilon_{k+1}}{\theta}} \leq \sqrt{\frac{\eta + \gamma}{\theta}} \leq \delta - \rho$$

De modo que:

$$\|x^{k+1} - x^*\| \leq \|x^{k+1} - x^k\| + \|x^k - x^*\| < \delta$$

■

Teorema 3.3.2 Sea $g: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia, semicontinua inferior y ε -Kurdyka Lojasiewicz. Sea $\{x^k\}$ una sucesión originada por el Algoritmo 10. Si $\{x^k\}$ es g -precompacta, entonces tiene longitud finita y $\{g(x^k)\}$ converge a un ε punto crítico.

Demostración:

Sea x^* el punto de acumulación específico de la Definición 1.1.17 i.e. $x^{k_j} \rightarrow x^*$ y $g(x^{k_j}) \rightarrow g(x^*)$. Sabemos por la condición 3.18 del Algoritmo 10 que la sucesión $\{g(x^k)\}$ es no creciente y por la Proposición 3.0.2 (i), sabemos que la sucesión converge. Por tanto, es seguro afirmar que $g(x^k) \rightarrow g(x^*)$ con $g(x^k) \geq g(x^*) \forall k \in \mathbb{N}$.

Ahora, por definición de límite, para cualquier $\eta > 0$, existe $N_0 \in \mathbb{N}$ tal que $|g(x^k) - g(x^*)| < \eta \forall k \geq N$. Así tenemos nuestra primera conclusión: a partir de un N lo suficientemente grande $g(x^*) \leq g(x^k) < g(x^*) + \eta \forall k \geq N$. Además, por Condición 3.17, existe $N_0 \in \mathbb{N}$ tal que para todo $k \geq N_0$, $\varepsilon_{k+1} < \gamma$.

Ahora, como g es ϵ -KL garantizamos la existencia de U, η, γ y $\varphi : [0, \eta] \rightarrow \mathbb{R}_+$ tales como los que aparecen en la Definición 1.4.4. Tomando δ y ρ tal que $0 < \rho < \delta$, $B(x^*, \delta) \subset U$ y de ser necesario $\sqrt{\frac{\eta+\gamma}{\theta}} \leq \delta - \rho$. Obtenemos nuestra segunda conclusión gracias al Lema 3.3.1, el cual es: $\forall k \in \mathbb{N}, x^k \in B(x^*, \rho) \implies x^{k+1} \in B(x^*, \delta)$ con $g(x^{k+1}) \geq g(x^*)$.

Por Condición 3.17, hacemos $A = 3\sqrt{\theta} - \kappa$, de modo que con lo anterior y la condición de continuidad de φ , a partir de un $N_1 \in \mathbb{N}$ lo suficientemente grande obtenemos lo siguiente:

$$\begin{aligned} & \|x^* - x^{N_1}\| + \left(\frac{\kappa}{A} + 1\right) \sqrt{\frac{g(x^{N_1}) + \epsilon_{N_1+1} - g(x^*)}{\theta}} \\ & + \frac{1}{A} \varphi(g(x^{N_1+1}) - g(x^*)) + \frac{1}{A} \sum_{k=N_1+1}^{N+N_1} \beta_k + \frac{3}{A} \sum_{k=N_1+1}^{N+N_1} \sqrt{\epsilon_{k+1}} < \rho \end{aligned}$$

Resumiendo, para un $K_0 > \max\{N, N_0, N_1\}$ obtenemos:

$$\text{H1 } g(x^*) \leq g(x^k) < g(x^*) + \eta \quad \forall k \geq K_0.$$

H2

$$\begin{aligned} & \|x^* - x^{K_0}\| + \left(\frac{\kappa}{A} + 1\right) \sqrt{\frac{g(x^{K_0}) + \epsilon_{K_0+1} - g(x^*)}{\theta}} \\ & + \frac{1}{A} \varphi(g(x^{K_0+1}) - g(x^*)) + \frac{1}{A} \sum_{k=K_0+1}^N \beta_k + \frac{3}{A} \sum_{k=K_0+1}^N \sqrt{\epsilon_{k+1}} < \rho \end{aligned}$$

donde $A = 3\sqrt{\theta} - \kappa$.

$$\text{H3 } \forall k \in \mathbb{N}, x^k \in B(x^*, \rho) \implies x^{k+1} \in B(x^*, \delta) \text{ con } g(x^{k+1}) \geq g(x^*)$$

A partir de este momento, $k \geq K_0$.

Afirmamos: Si $x^k \in B(x^*, \rho)$ y $f(x^*) < f(x^k)$ entonces

$$\|x^{k+1} - x^k\| \leq \sqrt{\frac{g(x^k) + \epsilon_{k+1} - g(x^*)}{\theta}} \quad (3.21)$$

En efecto, gracias a H3, $x^k \in B(x^*, \rho)$ implica $x^{k+1} \in B(x^*, \delta)$ con $g(x^{k+1}) \geq g(x^*)$; luego:

$$\|x^{k+1} - x^k\| \leq \sqrt{\frac{g(x^k) + \varepsilon_{k+1} - g(x^{k+1})}{\theta}} \leq \sqrt{\frac{g(x^k) + \varepsilon_{k+1} - g(x^*)}{\theta}}$$

la primera desigualdad viene de la Condición 3.18 del Algoritmo 10.

Afirmamos: Si $x^k \in B(x^*, \rho)$ entonces

$$\sum_{k=K_0+1}^{K_0+N} \|x^{k+1} - x^k\| \leq$$

$$\frac{\kappa}{A} \sqrt{\frac{g(x^{K_0}) + \varepsilon_{K_0+1} - g(x^*)}{\theta}} + \frac{1}{A} \varphi(g(x^{K_0+1}) - g(x^*)) + \frac{1}{A} \sum_{k=K_0+1}^{K_0+N} \beta_k + \frac{3}{A} \sum_{k=K_0+1}^{K_0+N} \sqrt{\varepsilon_{k+1}} \quad (3.22)$$

donde $A = 3\sqrt{\theta} - \kappa$

En efecto, sea $\Gamma^k = g(x^k) - g(x^*)$. Ya que φ es cóncava, entonces:

$$\varphi(\Gamma^{k+1}) - \varphi(\Gamma^k) \leq \varphi'(\Gamma^k)(\Gamma^{k+1} - \Gamma^k)$$

$$\varphi(\Gamma^k) - \varphi(\Gamma^{k+1}) \geq \varphi'(\Gamma^k)(\Gamma^k - \Gamma^{k+1})$$

$$\varphi(\Gamma^k) - \varphi(\Gamma^{k+1}) \geq \varphi'(\Gamma^k)(g(x^k) - g(x^{k+1})) \quad (3.23)$$

Por (3.18) tenemos:

$$\theta \|x^{k+1} - x^k\|^2 - \varepsilon_{k+1} \leq g(x^k) - g(x^{k+1})$$

Reemplazando en (3.23) obtenemos:

$$\varphi(\Gamma^k) - \varphi(\Gamma^{k+1}) \geq \varphi'(\Gamma^k)[\theta \|x^{k+1} - x^k\|^2 - \varepsilon_{k+1}] \quad (3.24)$$

Ahora como $x^k \in B(x^*, \rho)$ implica por H3 y Condición 3.18 que $g(x^*) \leq g(x^{k+1}) < g(x^k) <$

$g(x^*) + \eta$ podemos aplicar la desigualdad de épsilon-Kurdyka Lojasiewicz:

Si $v^k \in \partial_{\varepsilon_k} g(x^k)$ entonces $\varphi'(g(x^k) - g(x^*)) \|v^k\| \geq 1$ lo cual implica que $\|v^k\| \neq 0$ luego por Condición 3.20 tenemos $0 < \|v^k\| \leq \kappa \|x^k - x^{k-1}\| + \beta_k$, entonces:

$$\varphi'(\Gamma^k) = \varphi'(g(x^k) - g(x^*)) \geq \frac{1}{\|v^k\|} \geq \frac{1}{\kappa \|x^k - x^{k-1}\| + \beta_k}$$

Reemplazando en la Desigualdad 3.24, se obtiene:

$$\varphi(\Gamma^k) - \varphi(\Gamma^{k+1}) \geq \frac{\theta \|x^{k+1} - x^k\|^2 - \varepsilon_{k+1}}{\kappa \|x^k - x^{k-1}\| + \beta_k}$$

$$[\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})][\kappa \|x^k - x^{k-1}\| + \beta_k] \geq \theta \|x^{k+1} - x^k\|^2 - \varepsilon_{k+1}$$

Pueden ocurrir dos casos:

- CASO I: $0 \leq \theta \|x^{k+1} - x^k\|^2 - \varepsilon_{k+1}$

Usando la conocida desigualdad $2\sqrt{ab} \leq a + b$:

$$[\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + [\kappa \|x^k - x^{k-1}\| + \beta_k] \geq 2\sqrt{\theta \|x^{k+1} - x^k\|^2 - \varepsilon_{k+1}}$$

$$[\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + [\kappa \|x^k - x^{k-1}\| + \beta_k] \geq 2\sqrt{\theta} \|x^{k+1} - x^k\| - 2\sqrt{\varepsilon_{k+1}}$$

$$[\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + \kappa \|x^k - x^{k-1}\| + \beta_k + 2\sqrt{\varepsilon_{k+1}} \geq 2\sqrt{\theta} \|x^{k+1} - x^k\|$$

$$2\sqrt{\theta} \|x^{k+1} - x^k\| \leq \kappa \|x^k - x^{k-1}\| + [\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + \beta_k + 2\sqrt{\varepsilon_{k+1}} \quad (3.25)$$

-CASO II: $\theta \|x^{k+1} - x^k\|^2 - \varepsilon_{k+1} \leq 0$

entonces:

$$\sqrt{\theta} \|x^{k+1} - x^k\| \leq \sqrt{\varepsilon_{k+1}} \quad (3.26)$$

de ambos casos, (3.25) y (3.26):

$$3\sqrt{\theta}\|x^{k+1} - x^k\| \leq \kappa\|x^k - x^{k-1}\| + [\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + \beta_k + 3\sqrt{\varepsilon_{k+1}} \quad (3.27)$$

Ahora, sumando desde $k = K_0 + 1$ a $k = K_0 + N$:

$$\begin{aligned} & 3\sqrt{\theta} \sum_{k=K_0+1}^{K_0+N} \|x^{k+1} - x^k\| \leq \\ & \kappa \sum_{k=K_0+1}^{K_0+N} \|x^k - x^{k-1}\| + \sum_{k=K_0+1}^{K_0+N} [\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + \sum_{k=K_0+1}^{K_0+N} \beta_k + 3 \sum_{k=K_0+1}^{K_0+N} \sqrt{\varepsilon_{k+1}} \end{aligned}$$

Después de algunas operaciones aritméticas, tenemos:

$$\begin{aligned} (3\sqrt{\theta} - \kappa) \sum_{k=K_0+1}^{K_0+N-1} \|x^{k+1} - x^k\| & \leq \kappa[\|x^{K_0+1} - x^{K_0}\| - 3\sqrt{\theta}\|x^{K_0+N+1} - x^{K_0+N}\|] \\ & + \sum_{k=K_0+1}^{K_0+N} [\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + \sum_{k=K_0+1}^{K_0+N} \beta_k + 3 \sum_{k=K_0+1}^{K_0+N} \sqrt{\varepsilon_{k+1}} \end{aligned}$$

Dividiendo todo entre $A = 3\sqrt{\theta} - \kappa$

$$\begin{aligned} \sum_{k=K_0+1}^{K_0+N-1} \|x^{k+1} - x^k\| & \leq \frac{\kappa}{3\sqrt{\theta} - \kappa} \|x^{K_0+1} - x^{K_0}\| + \frac{1}{3\sqrt{\theta} - \kappa} \sum_{k=K_0+1}^{K_0+N} [\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] \\ & + \frac{1}{3\sqrt{\theta} - \kappa} \sum_{k=K_0+1}^{K_0+N} \beta_k + \frac{3}{3\sqrt{\theta} - \kappa} \sum_{k=K_0+1}^{K_0+N} \sqrt{\varepsilon_{k+1}} \end{aligned}$$

La desigualdad queda:

$$\begin{aligned} & \sum_{k=K_0+1}^{K_0+N-1} \|x^{k+1} - x^k\| \leq \\ & \frac{\kappa}{A} \|x^{K_0+1} - x^{K_0}\| + \frac{1}{A} \sum_{k=K_0+1}^{K_0+N} [\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + \frac{1}{A} \sum_{k=K_0+1}^{K_0+N} \beta_k + \frac{3}{A} \sum_{k=K_0+1}^{K_0+N} \sqrt{\varepsilon_{k+1}} \quad (3.28) \end{aligned}$$

$$\sum_{k=K_0+1}^{K_0+N+1} \|x^{k+1} - x^k\| \leq$$

$$\frac{\kappa}{A} \sqrt{\frac{g(x^{K_0}) + \varepsilon_{K_0+1} - g(x^*)}{\theta}} + \frac{1}{A} [\varphi(\Gamma^{K_0+1}) - \varphi(\Gamma^{K_0+N+1})] + \frac{1}{A} \sum_{k=K_0+1}^{K_0+N} \beta_k + \frac{3}{A} \sum_{k=K_0+1}^{K_0+N} \sqrt{\varepsilon_{k+1}}$$

$$\sum_{k=K_0+1}^{K_0+N-1} \|x^{k+1} - x^k\| \leq$$

$$\frac{\kappa}{A} \sqrt{\frac{g(x^{K_0}) + \varepsilon_{K_0+1} - g(x^*)}{\theta}} + \frac{1}{A} \varphi(g(x^{K_0+1}) - g(x^*)) + \frac{1}{A} \sum_{k=K_0+1}^{K_0+N} \beta_k + \frac{3}{A} \sum_{k=K_0+1}^{K_0+N} \sqrt{\varepsilon_{k+1}}$$

la penúltima desigualdad viene de 3.21 y con ello queda probado la desigualdad 3.22.

Ahora, probemos por inducción que $x^k \in B(x^*, \rho) \forall k \in \mathbb{N}$.

De H1 y H2 obtenemos que: $x^{K_0} \in B(x^*, \rho)$ y $g(x^*) \leq g(x^{K_0}) < g(x^*) + \eta$.

Para $k = K_0 + 1$:

$$\|x^* - x^{K_0+1}\| \leq \|x^* - x^{K_0}\| + \|x^{K_0} - x^{K_0+1}\| \leq \|x^* - x^{K_0}\| + \sqrt{\frac{g(x^{K_0}) + \varepsilon_{K_0+1} - g(x^*)}{\theta}} < \rho < \delta$$

La primera desigualdad viene por la desigualdad triangular, la segunda por la desigualdad 3.21, la tercera por H2.

Así tenemos que $x^{K_0+1} \in B(x^*, \rho)$.

Ahora, supongamos por hipótesis inductiva que $x^{K_0+N} \in (x^*, \rho)$ para algún N natural, luego la Desigualdad 3.22 se cumple.

Ahora, probemos que se cumple para x^{K_0+N+1}

$$\|x^* - x^{K_0+N+1}\| \leq \|x^* - x^{K_0}\| + \|x^{K_0+1} - x^{K_0}\| + \dots + \|x^{K_0+N+1} - x^{K_0+N}\|$$

$$\|x^* - x^{K_0+N+1}\| \leq \|x^* - x^{K_0}\| + \|x^{K_0+1} - x^{K_0}\| + \sum_{k=K_0+1}^{K_0+N-1} \|x^{k+1} - x^k\|$$

$$\begin{aligned} \|x^* - x^{K_0+N+1}\| &\leq \|x^* - x^{K_0}\| + \left(\frac{\kappa}{A} + 1\right) \sqrt{\frac{g(x^{K_0}) + \varepsilon_{K_0+1} - g(x^*)}{\theta}} \\ &+ \frac{1}{A} \varphi(g(x^{K_0+1}) - g(x^*)) + \frac{1}{A} \sum_{k=K_0+1}^N \beta_k + \frac{3}{A} \sum_{k=K_0+1}^N \sqrt{\varepsilon_{k+1}} < \rho \end{aligned}$$

La penúltima desigualdad viene de 3.22 y la última viene de H2.

Así tenemos $x^{N+1} \in B(x^*, \rho)$. Por inducción se concluye que $x^k \in B(x^*, \rho) \forall k \in \mathbb{N}$.

Por otro lado, haciendo $N \rightarrow +\infty$ en la Desigualdad 3.22, tenemos:

$$\begin{aligned} \sum_{k=K_0+1}^{\infty} \|x^{k+1} - x^k\| &\leq \\ \frac{\kappa}{A} \sqrt{\frac{g(x^{K_0}) + \varepsilon_{K_0+1} - g(x^*)}{\theta}} &+ \frac{1}{A} \varphi(g(x^{K_0+1}) - g(x^*)) + \frac{1}{A} \sum_{k=K_0+1}^{\infty} \beta_k + \frac{3}{A} \sum_{k=K_0+1}^{\infty} \sqrt{\varepsilon_{k+1}} \end{aligned}$$

Luego:

$$\sum_{k=K_0+1}^{+\infty} \|x^{k+1} - x^k\| \leq +\infty$$

Haciendo $x^k = x^{K_0+k}$ tenemos que x^k tiene longitud finita, lo que implica que $x^k \rightarrow x^*$.

Además, al inicio de esta demostración hemos concluido que $g(x^k) \rightarrow g(x^*)$. Luego x^k es un épsilon punto crítico de g .

En efecto, sean $v^k \in \partial_{\varepsilon_k} g(x^k)$ entonces $\|v^k\| \rightarrow 0$, pues $\|x^{k-1} - x^k\| \rightarrow 0$, $\beta_{k+1} \rightarrow 0$ y $\|v^k\| \leq \kappa \|x^{k-1} - x^k\| + \beta_{k+1}$

■

Capítulo IV

Velocidad de convergencia bajo las condiciones de la desigualdad ε -Kurdyka Lojasiewicz

En este capítulo estudiaremos la tasa de convergencia del Algoritmo Abstracto de ε -descenso generalizado para funciones ε -Kurdyka Lojasiewicz. Para ello, supongamos que $x^k \xrightarrow{g} x^*$ y g satisface la desigualdad de ε -Kurdyka Lojasiewicz en x^* , conservamos los objetos de φ , U , η , γ de la Definición 1.4.4. De esta manera planteamos el siguiente teorema:

Teorema 4.0.1 $\|x^K - x^*\| = O\{\max\{\sqrt{\Gamma^K + \gamma}, \varphi(\Gamma^K), 1\}\}$

Recordemos la Desigualdad 3.27.

$$3\sqrt{\theta}\|x^{k+1} - x^k\| \leq \kappa\|x^k - x^{k-1}\| + [\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + \beta_k + 3\sqrt{\varepsilon_{k+1}}$$

Sumando la desigualdad para $k = K, \dots, N$, para ello recordemos la Desigualdad 3.28:

$$\begin{aligned}
& \sum_{k=K+1}^{N-1} \|x^{k+1} - x^k\| \leq \\
& \frac{\kappa}{A} \|x^{K+1} - x^K\| + \frac{1}{A} \sum_{k=K+1}^N [\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + \frac{1}{A} \sum_{k=K+1}^N \beta_k + \frac{3}{A} \sum_{k=K+1}^N \sqrt{\varepsilon_{k+1}} \\
& \sum_{k=K}^{N-1} \|x^{k+1} - x^k\| \leq \\
& \frac{\kappa+A}{A} \|x^{K+1} - x^K\| + \frac{1}{A} \sum_{k=K+1}^N [\varphi(\Gamma^k) - \varphi(\Gamma^{k+1})] + \frac{1}{A} \sum_{k=K+1}^N \beta_k + \frac{3}{A} \sum_{k=K+1}^N \sqrt{\varepsilon_{k+1}} \\
& \sum_{k=K}^{N-1} \|x^{k+1} - x^k\| \leq \\
& \frac{\kappa+A}{A} \|x^{K+1} - x^K\| + \frac{1}{A} \varphi(\Gamma^{K+1}) + \frac{1}{A} \sum_{k=K+1}^N \beta_k + \frac{3}{A} \sum_{k=K+1}^N \sqrt{\varepsilon_{k+1}}
\end{aligned}$$

donde $A = 3\sqrt{\theta} - \kappa$ y $\Gamma^k = (g(x^k) - g(x^*))$. Usando la desigualdad triangular y pasando al límite tenemos:

$$\begin{aligned}
\|x^K - x^*\| & \leq \sum_{k=K}^{\infty} \|x^{k+1} - x^k\| \\
& \leq \frac{\kappa+A}{A} \|x^{K+1} - x^K\| + \frac{1}{A} \varphi(\Gamma^{K+1}) + \frac{1}{A} \sum_{k=K+1}^{\infty} \beta_k + \frac{3}{A} \sum_{k=K+1}^{\infty} \sqrt{\varepsilon_{k+1}} \\
& \leq \frac{\kappa+A}{A} \sqrt{\frac{g(x^K) + \gamma - g(x^*)}{\theta}} + \frac{1}{A} \varphi(\Gamma^{K+1}) + \beta + \alpha
\end{aligned}$$

donde $\frac{1}{A} \sum_{k=K+1}^{\infty} \beta_k = \beta < \infty$ y $\frac{3}{A} \sum_{k=K+1}^{\infty} \sqrt{\varepsilon_{k+1}} = \alpha < \infty$

Así tenemos:

$$\|x^K - x^*\| \leq \frac{\kappa+A}{A} \sqrt{\frac{g(x^K) + \gamma - g(x^*)}{\theta}} + \frac{1}{A} \varphi(\Gamma^{K+1}) + \beta + \alpha$$

$$\|x^K - x^*\| \leq \frac{\kappa + A}{A\sqrt{\theta}} \sqrt{\Gamma^k + \gamma} + \frac{1}{A} \varphi(\Gamma^{K+1}) + \beta + \alpha$$

Como Γ es decreciente y $g(x^*) \leq g(x^K)$ obtenemos:

$$\|x^K - x^*\| \leq \frac{\kappa + A}{A\sqrt{\theta}} \sqrt{\Gamma^k + \gamma} + \frac{1}{A} \varphi(\Gamma^K) + \beta + \alpha$$

Luego:

$$\|x^K - x^*\| = O\{\max\{\sqrt{\Gamma^k + \gamma}, \varphi(\Gamma^K), 1\}\}$$

Capítulo V

Ejemplos computacionales

En esta sección se presentaran algunos resultados computacionales con la intención de evidenciar que estos algoritmos de descenso que consideran errores abordados en la sección 2.2 de este trabajo, convergen. Debemos recalcar el hecho de que no se ha realizado un estudio de los criterios de error y hemos adoptado aquellos acondicionados para algoritmos exactos. Una cuestión que es tentador de estudiar y de ese modo disminuir incluso el número de iteraciones para obtener resultados aceptables.

Antes de seguir con nuestro cometido informamos que para los siguientes resultados fueron obtenidos a partir de programas escritos en el programa MATLAB R2023b (programas detallados en los anexos) en una laptop con las siguientes características:

- Procesador: 11th Gen Intel(R) Core(TM) i5-1155G7 @ 2.50GHz 2.50 GHz.
- RAM instalada: 12.0 GB (11.8 GB usable).
- Sistema: Sistema operativo de 64 bits, procesador basado en x64.

Método de gradiente inexacto con errores

El problema a resolver es el siguiente:

$$\text{mín}\{5x^2 + y^2, (x,y) \in \mathbb{R}^2\}$$

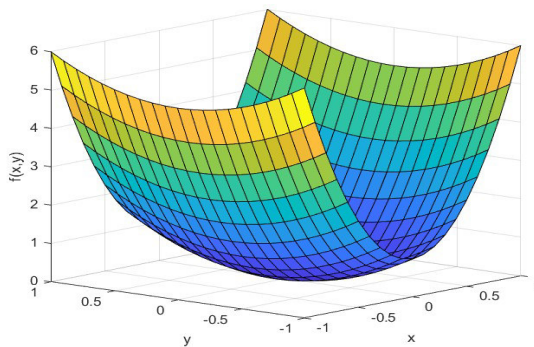


Figura 5.1: Gráfica de la función objetivo

Es claro que el problema tiene solución y alcanza mínimo en el punto $(0,0)$. Notamos que la función objetivo cumple con las condiciones requeridas para aplicar el método pues es una función de clase C^1 con gradiente 10-Lipschitz.

Con las condiciones cumplidas aplicaremos el método iterativo explicado en la Observación 2.2.2, para lo cual definimos los siguientes datos de entrada:

- Punto inicial: $x^0 = (2,2)$
- La constante de Lipschitz: $L = 10$
- La constante θ debe ser mayor a L : $\theta = 10,5$
- La constante κ debe ser mayor a θ : $\kappa = 11$. (Observación 2.2.2)

- El error escalar para la Condición 2.6: $\epsilon^{k+1} = 0,5^k$
- El error escalar para la Condición 2.7: $\beta^{k+1} = 0,5^{2k}$
- El tamaño de paso γ_k se elegiran de manera aleatoria en el rango de $[a, b]$. Donde $a = 1/\kappa$ y $b = 1/\theta$ (según la Ecuación 2.11 sabemos que $\frac{1}{2b} = \frac{\theta}{2}$ y por la Ecuación 2.12 $\frac{1}{a} = \kappa$).
- Del mismo modo, para aplicar este método iterativo 2.8 necesitamos el vector de error e , el cual conseguimos mediante un vector random que cumple las condiciones:
 - $\epsilon_{k+1} = \frac{\|e^k\|^2}{2a}$.
 - $\beta_{k+1} = \frac{1}{a}\|e^k\|$

Tabla 1: Resultados del método de gradiente inexacto con errores

Iter	$\nabla f(x^k)$	$\ x^{k+1} - x^k\ $	$x^k = (x_1^k, x_2^k)$	$f(x^k)$	ϵ_k	β_k
1	5.48021	1.59947	(0.412811 , 1.80218)	4.09993	0.5	3.31662
2	3.41951	0.372508	(0.059724 , 1.68347)	2.85192	0.25	2.34521
3	2.94959	0.217842	(0.028543 , 1.46788)	2.15873	0.0625	1.1726
4	2.50709	0.225958	(0.0339742 , 1.24198)	1.54829	0.015625	0.586302
5	2.05958	0.220147	(0.0252688 , 1.02201)	1.04769	0.00390625	0.293151
...	(... , ...)
50	0.00019	2.33e-05	(7.71e-16 , 9.97e-05)	9.39e-09	3.16e-30	8.33e-14
51	0.00016	1.93e-05	(3.43e-16 , 8.14e-05)	6.63e-09	7.89e-31	4.17e-15
52	0.00013	1.53e-05	(9.17e-17 , 6.61e-05)	4.37e-09	1.97e-31	2.08e-15
53	0.00011	1.20e-05	(7.92e-17 , 5.41e-05)	2.92e-09	4.93e-32	1.04e-15
54	8.78e-05	1.02e-05	(3.65e-17 , 4.39e-05)	1.93e-09	1.23e-32	5.21e-16

En la Tabla 1 podemos ver los resultados obtenidos y es evidente el acercamiento al mínimo de la función.

Método proximal inexacto

El problema a resolver es el siguiente:

$$\min\{g(x,y) = (x-2y)^2 + (x-2)^2 \text{ con } (x,y) \in \mathbb{R}^2\}$$

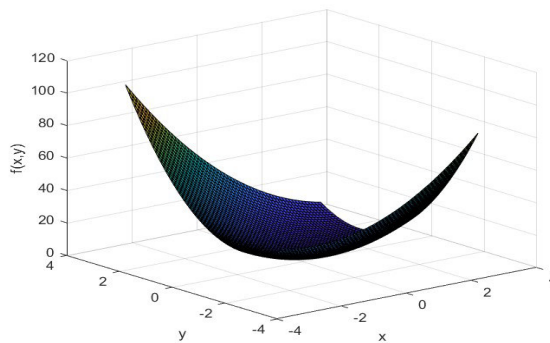


Figura 5.2: Gráfica de la función objetivo

Es claro que el problema tiene solución y alcanza mínimo en el punto $(2,1)$. Notamos que la función objetivo cumple con las condiciones requeridas para aplicar el Algoritmo 3 pues es una función continua y acotada desde abajo.

Vamos a aplicar Algoritmo 3 y para ello definimos los siguientes datos de entrada

- Punto inicial: $x^0 = (0,3)$
- La constante $0 \leq \omega < 1$: $\omega = 0,99$
- Para $\tau_k \in [\underline{\tau}, \bar{\tau}]$ con $0 < \underline{\tau} \leq \bar{\tau} < \infty$: $\underline{\tau} = 1$ y $\bar{\tau} = 2$
- La constante δ debe cumplir $0 < \delta \leq 1$: $\delta = 1$.
- El error escalar para la Condición 2.15: $\epsilon^{k+1} = 0,1^k$

Antes de mostrar los resultados debemos mencionar que primero nos enfrentamos al problema de encontrar un ε -mínimo dado por la Condición 2.15, es decir:

$$\min\{g(x,y) = (x-2y)^2 + (x-2)^2 + \frac{1}{2\tau_k} \|(x^k, y^k) - (x,y)\|^2 : (x,y) \in \mathbb{R}^2\}$$

Para el cual sabemos que existe solución y aproximamos mediante el Método de gradiente inexacto con búsqueda de Armijo, hasta conseguir un punto que cumpla con las Condiciones 2.16 y 2.17.

A continuación mostramos los resultados obtenidos, donde se puede ver un claro acercamiento al punto mínimo de la función.

Tabla 2: Resultados del método proximal inexacto con errores

Iter	$\nabla f(x^k)$	$\ x^{k+1} - x^k\ $	$x^k = (x_1^k, x_2^k)$	$f(x^k)$	ε_k
1	1.51624	2.64947	(1.75565 , 1.01572)	0.135759	0.1
2	0.173517	0.165591	(1.90323 , 0.940606)	0.00984928	0.264947
3	0.0418656	0.0864268	(1.97663 , 0.986233)	0.000563453	0.0165591
4	0.0135019	0.0187443	(1.99307 , 0.995244)	5.4717e-05	0.00864268
5	0.00388978	0.00602352	(1.99804 , 0.998637)	4.41882e-06	0.00187443
6	0.00130969	0.00160636	(1.99937 , 0.999547)	4.74533e-07	0.000602352
7	0.000389387	0.000527001	(1.99978 , 0.999872)	4.86796e-08	0.000160636
8	0.000134748	0.000165438	(1.99993 , 0.999957)	5.76104e-09	5.27001e-05
9	3.46405e-05	6.44615e-05	(1.99998 , 0.999989)	3.77974e-10	1.65438e-05
10	9.36943e-06	1.66832e-05	(2 , 0.999997)	2.40822e-11	6.44615e-06
11	2.82063e-06	3.7222e-06	(2 , 0.999999)	2.55095e-12	1.66832e-06
12	9.35557e-07	1.22327e-06	(2 , 1)	2.77552e-13	3.7222e-07

En la Tabla 2 es evidente que encontramos exactamente el punto mínimo de la función objetivo. Sin embargo; es necesario recalcar que en este trabajo solo garantizamos teóricamente la convergencia a puntos críticos.

Método inexacto de división hacia adelante y hacia atrás

Presentamos un problema estructurado denominando "Detección de compresión", el cual entre otras formas puede definirse de la siguiente manera:

Fijando un parámetro $\lambda > 0$:

$$\min\{\lambda\|x\|_0 + \frac{1}{2}\|Ax - b\|^2 \text{ donde } x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n} \text{ y } b \in \mathbb{R}^m\}$$

Donde $\|\cdot\|_0$ es la norma ℓ^0 que indica la cantidad de componentes diferentes de cero del vector.

Tenemos $\phi(x) = \lambda\|x\|_0$ como la función semicontinua inferior no convexa y no diferenciable y $\varphi(x) = \frac{1}{2}\|Ax - b\|^2$ como la función de clase C^1 con gradiente $\|A^T A\|_F$ -lipschitz.

Para ver que $\phi + \varphi$ es una función ε -KL, simplemente veamos que φ es una función polinómica y que $\|\cdot\|_0$ tiene una gráfica lineal por partes, por lo tanto, la suma $\phi + \varphi$ es semialgebraica. (Ver [33])

Tomamos $\varepsilon_k = 0,5^{(2k)}$ y como el subdiferencial de Fréchet para la función ϕ está definido por:

$$\partial\|x\|_0 = \{x^* \in \mathbb{R}^n : x_i^* = 0 \text{ para cada } i \text{ tal que } x_i \neq 0\}$$

aplicamos la Observación 2.2.3 para $m = 2, n = 3, A$ y b aleatorios para obtener los siguientes resultados:

Tabla 3: Resultados para el problema de detección de compresión

It	$\ w^k\ $	$\ x^{k+1} - x^k\ $	$x^k = (x_1^k, x_2^k, x_3^k)$	$f(x^k)$
1	0.108	0.080	(2.057,-1.943,0)	1.0039
2	0.036	0.015	(2.062,-1.929,0)	1.0035
3	0.027	0.014	(2.064,-1.916,0)	1.0031
4	0.022	0.012	(2.062,-1.904,0)	1.0029
...	(... , ... , ...)	...
149	0.00021	6.44e-05	(1.91761,-1.65332,0)	1
150	0.00020	0.0001	(1.91752,-1.6513,0)	1
151	0.00019	0.0001	(1.91746,-1.65307,0)	1
152	0.00018	0.0001	(1.9174,-1.65297,0)	1
153	0.00017	8.75e-06	(1.91739,-1.65296,0)	1

En la Tabla 3 es evidente que nos aproximamos a un punto crítico de la función objetivo; sin embargo, podemos notar que estos puntos siguen calculandose a pesar de ya haber alcanzado un valor mínimo de la función. Esto es porque a pesar de ser un algoritmo que considera errores le estamos exigiendo la misma exactitud que a cualquier algoritmo que no considera errores, por tal motivo un posible proyecto futuro podría estar enfocado en los criterios de error para este tipo de métodos.

Capítulo VI

Conclusiones

Presentamos un algoritmo abstracto de ϵ -descenso generalizado que cubre algoritmos de descenso inexactos para encontrar puntos críticos de una función semicontinua inferior y propia extendida, que tiene en cuenta errores escalares en cada iteración. Mostramos que si existe un punto de acumulación de la sucesión generada por el Algoritmo 1, este es un punto crítico límite generalizado. Además, bajo la condición de ϵ -Kurdyka Lojasiewicz se llega a la convergencia a un punto crítico.

Una contribución importante de nuestro trabajo se evidencia en la aplicación computacional de los métodos trabajados en el Capítulo IV, porque muchos algoritmos, como por ejemplo los métodos regularizados de gradiente inexacto, proximal, adelante-atrás y Gauss-Seidel, ya incluyen en cada aproximación errores escalares.

Observamos que la coercividad implica que la sucesión generada por el Algoritmo 1 está acotada, garantizando así la existencia de puntos de acumulación débiles (o puntos de acumulación en el caso de dimensiones finitas). De esta manera, se podría replantear los resultados del Capítulo IV aplicando algunos supuestos adicionales como semicontinuidad débilmente inferior de la función objetivo; sin embargo, para ello sería necesario investigar propiedades adicionales del ϵ -subdiferencial límite de Fréchet.

Este trabajo presenta resultados de convergencia débiles y también un análisis de la convergencia global del algoritmo 1 bajo la propiedad de Kurdyka-Lojasiewicz tal como fue estudiado por Attouch et al. [22] y Frankel et al. [19], para futuros trabajos quedaría pendiente estudiar la tasa de convergencia del algoritmo para saber si la introducción de errores escalares aumenta la velocidad de convergencia del algoritmo y también estudiar los criterios de parada con la intención de disminuir aún más el número de iteraciones.

Bibliografía

- [1] Absil P. A., Mahony R. y Andrews B. «Convergence of the iterates of descent methods for analytic cost functions». En: *SIAM Journal on Optimization* 16.2 (2005), págs. 531-547.
- [2] Bihain A. «Optimization of upper semidifferentiable functions». En: *Journal of Optimization Theory and Applications* 44 (1984), págs. 545-568.
- [3] Goldstein A. «Optimization of Lipschitz continuous functions». En: *Mathematical Programming* 13 (1977), págs. 14-22.
- [4] Jofre A., Luc D.T. y Thera M. « ϵ -subdiferential and ϵ -monotonicity». En: *Nonlinear Analysis, Theory, Methods and Applications* 33.1 (1998), págs. 71-90.
- [5] Papa Quiroz E. A., Soubeyran A. y Oliveira P. R. «Coercivity and generalized proximal algorithms: application—traveling around the world». En: *Annals of Operations Research* (2022). Accepted: 11 April 2022.
- [6] Soubeyran A. y Souza J. C. D. O. «General descent method using w-distance. application to emergence of habits following worthwhile moves». En: *Journal of Nonlinear and Variational Analysis* 4.2 (2020), págs. 285-300.
- [7] Andrews B. Absil P. Mahony R. «Convergence of the iterates of descent methods for analytic cost functions». En: *SIAM Journal on Optimization* 16.2 (2006), págs. 531-547.

- [8] Hedy Attouch et al. «Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-Łojasiewicz Inequality». En: *Math. Oper. Res.* 35 (2010), págs. 438-457.
- [9] Goode J. J. Bazaraa M. S. y Nashed v Z. Z. «On the cones of tangents with applications to mathematical programming». En: *Journal of Optimization Theory and Applications* 13 (1974), págs. 389-426.
- [10] Jérôme Bolte, Aris Daniilidis y Adrian Lewis. «The Łojasiewicz Inequality for Nonsmooth Subanalytic Functions with Applications to Subgradient Dynamical Systems». En: *Society for Industrial and Applied Mathematics* 17 (2007), págs. 1205-1223.
- [11] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Nov. de 2011. ISBN: 978-0-387-70913-0. DOI: 10.1007/978-0-387-70914-7.
- [12] Mordukhovich BS. «Extremal Principle in Variational Analysis». En: *Análisis variacional y diferenciación generalizada I. Grundlehren der mathematischen Wissenschaften* 330 (2006).
- [13] Annette Burden, Richard Burden y J. Faires. *Numerical Analysis, 10th ed.* Ene. de 2016. ISBN: 1305253663. DOI: 10.13140/2.1.4830.2406.
- [14] Bustos. « ϵ -gradients pour les fonctions localement lipschitziennes et applications». En: *Numerical Functional Analysis and Optimization* 15 (1994), págs. 435-453.
- [15] Bento G. C. et al. «Inexact Multi-Objective Local Search Proximal Algorithms: Application to Group Dynamic and Distributive Justice Problems». En: *Journal of Optimization Theory and Applications* 177.1 (2018), págs. 181-200.
- [16] de Carvalho B. G. et al. «A proximal point method for difference of convex functions in multi-objective optimization with application to group dynamic problems». En: *Computational Optimization and Applications* 75.1 (2020), págs. 263-290.

- [17] Asplund E. «Fréchet differentiability of convex functions». En: *Acta Math.* 121 (1968), págs. 31-47.
- [18] Oliveira P.R. Ferreira O.P. «Subgradient Algorithm on Riemannian Manifolds». En: *Journal of Optimization Theory and Applications* 97.1 (1998), págs. 93-104.
- [19] Peypouquet J. Frankel P. Garrigos G. «Splitting Methods with Variable Metric for Kurdyka - Lojasiewicz Functions and General Convergence Rates». En: *Journal of Optimization Theory and Applications* 3.165 (2015), págs. 874-900.
- [20] P. Gajardo. «Introducción al Análisis Convexo. Instituto de Matemáticas.s». En: *Pontificie Universidad Católica de Valparaíso* (2006).
- [21] Attouch H. y Soubeyran A. «Local Search Proximal Algorithms as Decision Dynamics with Costs to Move». En: *Set-Valued Anal* 19 (2011), págs. 157-177.
- [22] Attouch H., Bolte J. y Svaiter B. F. «Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods». En: *Mathematical Programming* 137 (2013), págs. 91-129.
- [23] Attouch H., Briceño-Arias L. M. y Combettes P. L. «A parallel splitting method for coupled monotone inclusions». En: *SIAM Journal on Control and Optimization* 48.5 (2010), págs. 3246-3270.
- [24] Curry H.B. «The method of steepest descent for non-linear minimization problems». En: *The Quarterly of Applied Mathematics* 2 (1944), págs. 258-261.
- [25] Jean-Baptiste Hiriart-Urruty y Claude Lemaréchal. *Approximate Subdifferentials of Convex Functions*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, págs. 91-136. ISBN: 978-3-662-06409-2. DOI: 10.1007/978-3-662-06409-2_3. URL: https://doi.org/10.1007/978-3-662-06409-2_3.

- [26] Ekeland I. y Lebourg G. «Sous-gradients approchés et applications». En: *Comptes Rendus de l'Académie des Sciences* 281 (1975), págs. 219-222.
- [27] Palis J. y Melo W. «Geometric theory of dynamical systems». En: *Springer, New York, Berlin* (1982).
- [28] Olivier Ley Jerome Bolte Aris Daniilidis y Laurent Mazet. «Characterizations of Lojasiewicz inequalities: Subgradient flows, talweg, convexity». En: *Transactions of the American Mathematical Society* 6.362 (2010), págs. 3319-3363.
- [29] A. Parusinski K. Kurdyka. «Wf-stratification of subanalytic functions and theLojasiewicz inequality». En: *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics* 318 (1994), págs. 129-133.
- [30] Buttazzo G. KAttouch H. y Michailen G. «Variational Analysis in Sobolev and BV Spaces». En: *MPS-SIAM Series on Optimization. Springer, NewYorks* (2014).
- [31] Erwin Kreyszing. *Introductory Functional Analysis with Applications*. Wiley classics library. Wiley India Pvt. Limited, 2007. ISBN: 9788126511914. URL: <https://books.google.com.pe/books?id=osXw-pRsptoC>.
- [32] A.Y. Kruger. «On Fréchet Subdifferentials». En: *Journal of Mathematical Sciences* 116 (2003), págs. 3325-3358.
- [33] K. Kurdyka. «On gradients of functions definable in o-minimal structures». En: *Annales de l'Institut Fourier* 48.3 (1998), págs. 769-783.
- [34] Combettes P. L. y Wajs V. R. «Signal recovery by proximal forward–backward splitting». En: *Multiscale Model. Simul.* 4 (2005), págs. 1168-1200.
- [35] Stanisław Łojasiewicz. «Ensembles semi-analytiques». En: *Inst. Hautes Etudes Sci., Bures-sur-Yvette, France* (1965), pág. 92.

- [36] Stanisław Łojasiewicz. «Une propriété topologique des sous-ensembles analytiques réels». En: *Les Équations aux Dérivées Partielles, Editions du Centre National de la Recherche Scientifique, Paris* (1963), págs. 87-89.
- [37] William Paulsen. *Asymptotic Analysis and Perturbation Theory*. 2013, pág. 550. ISBN: 9780429189531. DOI: <https://doi.org/10.1201/b15165>.
- [38] J. Peypouquet. *Optimización y Sistemas Dinámicos*. 2013. ISBN: 978-0-387-70913-0.
- [39] Mayne D. Q. Polak E. y Wardi Y. «On the extension of constrained optimization algorithms from differentiable to nondifferentiable problems». En: *SIAM Journal on Control and Optimization* 21 (1983), págs. 179-203.
- [40] Trojanski S.L. «On locally uniformly convex and differentiable norm in certain non-separable Banach spaces». En: *Studia Math.* 37 (1971), págs. 73-180.
- [41] Polyak B. T. «Introduction to optimization». En: *Optimization Software, New York*. (1987).
- [42] Rockafellar R. T. y Wets R. J. B. «Variational Analysis, Grundlehren der Mathematischen Wissenschaften». En: *Springer Berlin, Heidelberg* 317 (1998).
- [43] Bao T.Q. et al. «Vector Optimization with Domination Structures: Variational Principles and Applications». En: *Set-Valued and Variational Analysis* 30.2 (2022), págs. 695-729.
- [44] Oliveira P.R. TMoreno F.A. G y Soubeyran A. «A proximal algorithm with quasi distance». En: *Application to habit's formation. Optimization* 61 (2012), págs. 1383-1403.
- [45] Dixon L. C. W. «Reflections on nondifferentiable optimization, Part 1. Ball gradient». En: *Journal of Optimization Theory and Applications* 32.2 (1980), pág. 165 174.
- [46] Kruger A. Ya. y Mordukhovich B. «Extremal points and the Euler equation in nonsmooth optimization». En: *Doklady Akademii Nauk BSSR* 24 (1980), págs. 684-687.

Anexo

Algoritmo de gradiente inexacto con errores

```
syms x y
F=5*x^2+y^2;
f=inline(F);
G=gradient(F);
g=inline(G);
L=10;
x0=[2 2];
t=10.5;
e=0.5;
k=11;
a=1/k;
b=1/t;
parar=0.0001;
nmax=100;
contador=0;
```

```

paso=0;

while contador<=nmax
    contador=contador+1;
    norme=sqrt(e*2*a);
    ek=rand(2,1);
    ek=(ek/norm(ek))*norme;
    r = (b-a).*rand(1) + a;
    x1=x0-r*g(x0(1),x0(2)).'+ek.';
    v=g(x0(1),x0(2)).';
    v1=g(x1(1),x1(2)).';
    fprintf(' %6.0f %12.6g %12.6g %14.6g %14.6g %16.6g
           %16.6g %16.6g\n', contador, x1(1),x1(2),norm(v1),
           norm(x1-x0),f(x1(1),x1(2)),e,norme/a);
    if norm(v1)<=0.0001
        contador=nmax+10;
        disp('fin');
    end
    x0=x1;
    e=1/2^(2*contador);
end

```

Algoritmo proximal inexacto con errores

```
syms x y
F=(x-2)^2+(x-2*y)^2;
f=inline(F);
G=gradient(F);
g=inline(G);
e=0.1;
x0=[0 3];
w=0.9;
d=1;
t_min=1;
t_max=2;
contador=0;
Nmax=100;
while contador<=Nmax
    contador=contador+1;
    t_k=(t_max-t_min)*rand(1) + t_min;
    %solucionamos el prob: min{f(x)+d/(2*t_k)|x-x0|^2}
    x1=prox(f(x0,d,t_k,e));
    primer=f(x1(1),x1(2))+(d/(t_k*2))*norm(x1-x0)^2<=f(x0
        (1),x0(2))+e;
    z=rand(1,2);
    z=e/norm(z)*z;
    v=g(x1(1),x1(2)).'+z;
```

```

segunda=norm(t_k*v+x1-x0)^2 <= w*(norm(t_k*v)^2+norm(x1
-x0)^2);
while primer~=1 && segunda~=1
    disp('yes')
    x1=x1;
end
if norm(x1-x0) < 0.000001
    disp('fin')
    contador=Nmax+10;
else
    fprintf(' %6.0f %12.6g %12.6g %14.6g %14.6g %16.6g
%16.6g \n', contador, x1(1),x1(2),norm(g(x1(1),
x1(2))),norm(x1-x0),f(x1(1),x1(2)),e);
    e=0.1*norm(x1-x0);
    x0=x1;
end
end
end

```

```

function X = proxf(x0,d,t,e)
    syms x y
    F=(x-2)^2+(x-2*y)^2+d/(2*t)*((x-x0(1))^2+(y-x0(2))^2);
    f=inline(F);
    G=gradient(F);
    g=inline(G);
    s=1;

```



```

sigma=0.1;
beta=0.5;
k=1;
k_max=2;
while k<=k_max
    v=g(x0(1),x0(2)).';
    normv=norm(v);
    if normv >= e
        m=0;
        factual=f(x0(1),x0(2));
        pnuevo=x0-v;
        fnuevo=f(pnuevo(1), pnuevo(2));
        dif=factual-fnuevo;
        lambda=1;
        test=sigma*lambda*normv^2;
        while dif < test
            m=m+1;
            lambda=beta^m*s;
            ppnuevo=x0-lambda*v;
            dif=f(x0(1),x0(2))-f(ppnuevo(1), ppnuevo(2));
            test=sigma*lambda*normv^2;
        end
        x0=x0-lambda*v;
    else
        k=k_max+10;
    end
end

```

```

        X=x0;
    end
end
X=x0;
end

```

Algoritmo inexacto de división hacia adelante y hacia atrás

```

%% min {mu*|x|_0+1/2|Ax-b|^2}
% g(x)=mu*|x|_0 y f(x)=1/2|Ax-b|^2

% Pedir n y m
%n=input('Ingrese la dimension n: \n');
%m=input('Ingrese la dimension m: \n');
n=3;
m=2;
mu = 0.5;
fprintf('Division hacia adelante y hacia atras para
    Deteccion Comprimida\n');
disp('Error escalar en la condicion de descenso suficiente:
    0.5^(2k)')
disp('-----')

```

```

disp(' I      x      |Grad(f)+v|      |x_k-x_{k+1}|      f(x(1)
      ,x(2))      e ')
disp(' ----- ')

% Con las dimensiones, elegir los datos de entrada: x0, mu,
      A, b
x0 = randi([-3,4], n,1);
A=rand(m,n);
b=rand(m,1);

% Paso gradiente con longitud de paso en un intervalo de 1<
      gamma_k<1/L
% f(z) = 0.5 ||Ax - b||^2
f = @(x) 0.5*norm(A*x-b, 'fro')^2;
grad = @(x) A'*A*x-A'*b;
% g(z) = ||x||_inf
g = @(x) mu*nnz(x);
L= norm(A'*A, 'fro');
epsilon =0.5;

Kmax=3000;
k=1;

while k<=Kmax

```

```

gamma_k = (1/L-0.00001-0.0000000000001).*rand(1,1) +
    0.000000000001;
x11=x0-gamma_k*grad(x0);
x1=prox_implementacion(x11,mu,gamma_k,n);
v=eesubdiff(x1,n);
v1=norm(grad(x0)+v);
cond1= g(x1)+dot(x1-x0,grad(x0))+(L+1)/2*norm(x1-x0)^2
    < g(x0) + epsilon^(2*k);
cond2=norm(grad(x1)+v)<=100*norm(x1-x0);
if cond1 && cond2
    fprintf('%6.0f %12.6g %12.6g %12.6g %14.6g %14.6g
        %16.6g %16.6g \n', k, x1(1),x1(2), x1(3) ,norm(
            v1),norm(x1-x0),f(x1)+g(x1),epsilon^k);
    if norm(x1-x0)<0.00001
        k=Kmax+100;
    end
    x0=x1;
end
k=k+1;
end

```

```

function prox = prox_implementacion(x, mu, gamma,n)
m=sqrt(2*mu*gamma);
r=zeros(n,1);
for i=1:n

```

```
    if abs(x(i))>m
        r(i)=x(i);
    elseif abs(x(i))==m
        r(i)=x(i);
    else
        r(i)=0;
    end
end
prox=r;
end
```

```
function w = eesubdiff(x1,n)
    r = (0.00000001+0.00000001).*rand(n,1) -0.00000001;
    for i=1:n
        if x1(i)~=0
            r(i)=0;
        end
    end
end
w=r;
end
```
