

**XXVI ESCUELA VENEZOLANA DE MATEMÁTICAS
EMALCA-VENEZUELA 2013**

**OPTIMIZACIÓN Y SISTEMAS
DINÁMICOS**

Juan Peypouquet

MÉRIDA, VENEZUELA, 1 AL 6 DE SEPTIEMBRE DE 2013

XXVI ESCUELA VENEZOLANA DE MATEMÁTICAS
EMALCA - VENEZUELA 2013

OPTIMIZACIÓN Y SISTEMAS DINÁMICOS

Juan Peypouquet Urbaneja

Universidad Técnica Federico Santa María

juan.peypouquet@usm.cl

MÉRIDA, VENEZUELA, 1 AL 6 DE SEPTIEMBRE DE 2013

XXVI ESCUELA VENEZOLANA DE MATEMÁTICAS

La Escuela Venezolana de Matemáticas es una actividad de los postgrados en matemáticas de las instituciones siguientes: Centro de Estudios Avanzados del Instituto Venezolano de Investigaciones Científicas, Facultad de Ciencias de la Universidad Central de Venezuela, Facultad de Ciencias de la Universidad de Los Andes, Universidad Simón Bolívar, Universidad Centroccidental Lisandro Alvarado y Universidad de Oriente, y se realiza bajo el auspicio de la Asociación Matemática Venezolana. La XXVI Escuela Venezolana de Matemáticas recibió financiamiento de la Academia de Ciencias Físicas, Matemáticas y Naturales de Venezuela, el Banco Central de Venezuela, el Fondo Nacional de Ciencia, Tecnología e Innovación (FONACIT), el Instituto Venezolano de Investigaciones Científicas (Centro de Estudios Avanzados, Departamento de Matemáticas y Ediciones IVIC), la Universidad de los Andes (CEP, CD-CHT, Departamento de Matemáticas de la Facultad de Ciencias, Decanato de Ciencias y Vicerrectorado Administrativo), Unión Matemática de América Latina y el Caribe (UMALCA) y Centre International de Mathématiques Pures et Appliquées (CIMPA).

2010 Mathematics Subject Classification: 46N10, 47N10, 49M37, 90C25

©Ediciones IVIC

Instituto Venezolano de Investigaciones Científicas

Rif: G-20004206-0

Optimización y Sistemas Dinámicos

Juan Peypouquet Urbaneja

Diseño y edición: Escuela Venezolana de Matemáticas

Preprensa e impresión: Gráficas Lauki C. A.

Depósito legal If66020130012420

ISBN 978-980-261-142-3

Caracas, Venezuela

2013

A Stefania

Prefacio

Este libro de texto contiene una gama relativamente amplia de conceptos y resultados sobre optimización, y ha sido diseñado como un apoyo para el curso

Sistemas Dinámicos y Minimización Convexa en Espacios de Hilbert,

parte de la XXVI Escuela Venezolana de Matemáticas y EMALCA - Venezuela, 2013.

Si bien el curso se centra en la minimización de funciones convexas en espacios de Hilbert, una parte importante de este libro texto se sitúa en un contexto más general. En él abordamos tres grandes temas, a saber:

1. Existencia de solución. Comenzamos por presentar un resultado abstracto que garantizan la existencia de solución para el problema de minimización en espacios topológicos, para luego estudiar lo que sucede en el caso convexo en espacios reflexivos. Si bien nos enfocamos en el problema de minimización, el de maximización se puede tratar de manera muy similar, usando las ideas presentadas aquí.
2. Condiciones de optimalidad. Presentamos una condición necesaria de optimalidad, basada en la primera derivada, para el caso de funciones diferenciables definidas en un espacio vectorial normado. Esta condición se generaliza después para funciones convexas, no necesariamente diferenciables. Gracias a la convexidad, además de necesaria, la condición es suficiente.

3. Aproximación de las soluciones mediante sistemas dinámicos. Estudiamos algunos sistemas dinámicos, a tiempo continuo y discreto, que producen trayectorias o sucesiones que convergen a soluciones del problema de minimización convexa en espacios de Hilbert.

Agradecimientos

Quisiera aprovechar la ocasión para extender un reconocimiento a todas las personas e instituciones vinculadas con la organización de la Escuela Venezolana de Matemáticas por la gran labor que han llevado a cabo durante décadas. Además agradezco al Comité Científico de la XXVI Escuela Venezolana de Matemáticas y EMALCA - Venezuela, 2013, la invitación a dictar el curso y a preparar este libro de texto. Como matemático venezolano que reside desde hace muchos años en el exterior, es para mí un placer y un privilegio participar en esta actividad.

Finalmente, deseo agradecer a los diversos entes que han contribuido a mi investigación, ya sea mediante financiamiento¹ o a través de mi formación². Cada faceta de mi trabajo académico ha jugado un papel clave en dar forma a las ideas presentadas aquí.

Juan Peypouquet Urbaneja

¹Proyecto FONDECYT 11090023, CMM - Universidad de Chile, Núcleo Milenio Información y Coordinación en Redes ICM/FIC P10-024F, Proyecto Anillo ACT-1106: Análisis de Problemas de Control y Aplicaciones, Universidad Técnica Federico Santa María

²Universidad de Chile, Universidad Pedro y María Curie - Paris VI, y Universidad Simón Bolívar.

Índice general

Prefacio	v
1 Introducción	1
I El Problema de Minimización	5
2 Conceptos generales	7
2.1 Funciones en los Reales Extendidos	7
2.2 Convexidad	8
2.3 Minimización Local y Global	10
3 Existencia de Soluciones	13
3.1 El Teorema de Weierstrass-Hilbert-Tonelli	13
3.2 Breve Repaso de Análisis Funcional	15
3.3 Minimización Convexa	18
4 Optimalidad	21
4.1 Caso Diferenciable	21
4.2 Caso Convexo, no Diferenciable	23
4.2.1 Subdiferencial y Regla de Fermat Generalizada	23
4.2.2 Continuidad y Subdiferenciabilidad	27
5 Ejemplos	35
5.1 Operador de Proximidad	35
5.2 Cálculo de Variaciones	37
5.3 Algunas Ecuaciones Elípticas	39

II	Sistemas Dinámicos en Minimización Convexa	41
6	Sistemas Dinámicos	43
6.1	Aspectos Generales	43
6.2	Sistemas Dinámicos y Minimización	44
6.3	Análisis Asintótico	46
6.3.1	Convergencia Débil	47
6.3.2	Convergencia Fuerte	48
7	Método del Máximo Descenso	51
7.1	Propiedades del Sistema	52
7.2	Convergencia	53
8	Algoritmo del Punto Proximal	57
8.1	Propiedades del Sistema	58
8.2	Convergencia	59
8.3	Consistencia de las Direcciones	61
9	Algoritmo del Gradiente	63
9.1	Preliminares	64
9.2	Propiedades del Sistema	65
9.3	Convergencia	67
10	Comentarios Finales	69
10.1	Caso de una Función Cuadrática	69
10.1.1	Método del Máximo Descenso	70
10.1.2	Algoritmo del Punto Proximal	70
10.1.3	Algoritmo del Gradiente	71
10.2	Estrategias de Descomposición	72
A	Espacios de Hilbert	73
A.1	Conceptos Básicos	73
A.2	Proyección Ortogonal y Dualidad	76

Capítulo 1

Introducción

Sean X un conjunto, $C \subset X$ y $f : X \rightarrow \mathbf{R}$ una función. En este curso consideraremos el *problema de minimización con restricciones*, que consiste en encontrar, si existe, el menor valor que f toma en C (llamado *valor del problema*) y determinar en qué elementos de C alcanza dicho valor. Para referirnos a este problema solemos escribir:

$$(\mathcal{P}) \quad \begin{cases} \text{Minimizar} & f(x) \\ \text{Sujeto a} & x \in C, \end{cases}$$

o también

$$\text{mín}\{f(x) : x \in C\}.$$

Un punto $x^* \in C$ es un *minimizador* de f en C si $f(x^*) \leq f(x)$ para todo $x \in C$. El conjunto de los minimizadores se llama *argumento del mínimo* y lo denotamos indistintamente por $\text{argmin}\{f(x) : x \in C\}$, $\text{argmin}_C(f)$, o simplemente $\text{argmin}(f)$, si el *conjunto factible* C está sobrentendido. En este contexto nos referiremos a f como la *función objetivo*.

Surgen inmediatamente dos preguntas: ¿Cómo sabemos si una función tiene minimizadores? ¿Cómo los encontramos? En este curso intentaremos responder a estas preguntas en el caso en que C es un subconjunto de un espacio de Hilbert real y $f : C \rightarrow \mathbf{R}$ es una función convexa, propia e inferiormente semicontinua (daremos las definiciones oportunamente).

Comencemos por observar que si C es un subconjunto de \mathbf{R}^N , hay dos resultados elementales que se estudian en la mayoría de los cursos de

Cálculo en Varias Variables, y que permiten responder total o parcialmente estas preguntas. El primero tiene que ver con la existencia de minimizadores bajo hipótesis esencialmente topológicas:

Teorema 1.1 (Teorema de Weierstrass). *Si C es un subconjunto cerrado, acotado y no vacío de \mathbf{R}^N y $f : C \rightarrow \mathbf{R}$ es una función continua, entonces f alcanza su máximo y su mínimo en C .*

El segundo nos da una *condición necesaria de optimalidad*; es decir, una propiedad que deben satisfacer todos los minimizadores:

Teorema 1.2 (Regla de Fermat). *Sea A un subconjunto abierto de \mathbf{R}^N . Si $x^* \in A$ es un minimizador de $f : A \rightarrow \mathbf{R}$ en A y f es diferenciable en x^* , entonces $\nabla f(x^*) = 0$.*

La Regla de Fermat es una condición necesaria. Otros resultados, basados típicamente en la segunda derivada, entregan *condiciones suficientes de optimalidad*. Veremos más adelante que, en el caso convexo, una condición análoga a la que entrega el Teorema 1.2 es, en efecto, suficiente. Pospondremos las demostraciones del Teorema de Weierstrass y la Regla de Fermat, pues los presentaremos en versiones más generales. Sin embargo, los utilizaremos para estudiar en detalle un ejemplo sencillo y muy conocido, que nos permitirá presentar algunas de las ideas que desarrollaremos más adelante:

Ejemplo 1.3. Supongamos que se cuenta con una muestra $(x_k, y_k)_{k=1}^m$, obtenida mediante m mediciones experimentales de las variables x e y . Supongamos, además, que hay buenas razones para sospechar que entre dichas variables existe una relación lineal-afín: $y = ax + b$, donde los parámetros a y b son desconocidos. La *aproximación por mínimos cuadrados* es una herramienta estadística que consiste en determinar los valores de a y b que mejor se ajustan a la muestra, en el sentido de que el *error cuadrático medio*:

$$E(a, b) = \frac{1}{m} \sum_{k=1}^m (ax_k + b - y_k)^2$$

sea lo menor posible. Observemos que la función $E : \mathbf{R}^2 \rightarrow \mathbf{R}$ es infinitamente diferenciable (en particular, es continua).

Existencia de minimizadores: Si bien la función E es continua, su dominio \mathbf{R}^2 no es acotado. Por lo tanto, no se puede utilizar directamente el Teorema de Weierstrass. Para solventar esta situación, restringiremos la función E a un conjunto C cerrado y acotado que nos permita utilizar el Teorema. Primero notemos que

$$\inf\{E(a, b) : (a, b) \in \mathbf{R}^2\} \leq E(0, 0)$$

y escribamos

$$C = \{(a, b) \in \mathbf{R}^2 : E(a, b) \leq E(0, 0)\}.$$

Si denotamos por $E|_C$ la restricción de E a C , es claro que

$$\operatorname{argmin}(E) = \operatorname{argmin}(E|_C).$$

Como E es continua, C es cerrado. Por otra parte, tengamos en cuenta que $\lim_{\|(a,b)\| \rightarrow +\infty} E(a, b) = +\infty$. Luego, existe $N > 0$ tal que $E(a, b) > E(0, 0)$ si $\|(a, b)\| > N$. En otras palabras, existe $N > 0$ tal que si $E(a, b) \leq E(0, 0)$, entonces $\|(a, b)\| \leq N$. Esto nos dice que C es acotado. Concluimos entonces que $\operatorname{argmin}(E) \neq \emptyset$.

Condición necesaria de optimalidad: Si (a^*, b^*) es un minimizador de E , el Teorema 1.2 nos dice que

$$\begin{aligned} (0, 0) &= \nabla E(a^*, b^*) \\ &= \frac{2}{m} (a^* \|\mathbf{x}\|^2 + b^* (\mathbf{x} \cdot \mathbf{1}) - (\mathbf{x} \cdot \mathbf{y}), a^* (\mathbf{x} \cdot \mathbf{1}) + b^* m - (\mathbf{y} \cdot \mathbf{1})). \end{aligned}$$

Aquí el símbolo \cdot denota el producto escalar en \mathbf{R}^m y hemos escrito $\mathbf{x} = (x_1, \dots, x_m)$, $\mathbf{y} = (y_1, \dots, y_m)$ y $\mathbf{1} = (1, \dots, 1) \in \mathbf{R}^m$, para simplificar la notación.

Lo anterior es equivalente al sistema de ecuaciones

$$\begin{cases} \|\mathbf{x}\|^2 a^* + (\mathbf{x} \cdot \mathbf{1}) b^* = \mathbf{x} \cdot \mathbf{y} \\ (\mathbf{x} \cdot \mathbf{1}) a^* + m b^* = \mathbf{y} \cdot \mathbf{1}, \end{cases}$$

que tiene solución única (salvo si todos los x_k son iguales, lo cual no tiene sentido experimental) dada por

$$a^* = \frac{m(\mathbf{x} \cdot \mathbf{y}) - (\mathbf{x} \cdot \mathbf{1})(\mathbf{y} \cdot \mathbf{1})}{m\|\mathbf{x}\|^2 - (\mathbf{x} \cdot \mathbf{1})^2}, \quad b^* = \frac{(\mathbf{y} \cdot \mathbf{1})\|\mathbf{x}\|^2 - (\mathbf{x} \cdot \mathbf{1})(\mathbf{x} \cdot \mathbf{y})}{m\|\mathbf{x}\|^2 - (\mathbf{x} \cdot \mathbf{1})^2}.$$

Como E tiene al menos un minimizador y que hay un solo elemento de \mathbf{R}^2 que satisface la condición necesaria de optimalidad, concluimos que

$$\operatorname{argmin}\{E(a, b) : (a, b) \in \mathbf{R}^2\} = \{(a^*, b^*)\},$$

con lo cual hemos resuelto completamente el problema.

Organización del texto:

La Parte I corresponde al problema de minimización en un marco más bien abstracto. En el Capítulo 2 presentaremos los conceptos básicos relacionados con el problema de minimización en general, y de minimización convexa en particular. El Capítulo 3 contiene los resultados de existencia de solución para el problema de minimización, con o sin convexidad. En el Capítulo 4 discutiremos sobre condiciones de optimalidad bajo hipótesis de diferenciabilidad (condiciones necesarias) o de convexidad (condiciones necesarias y suficientes) de la función objetivo. Terminamos esta parte con el Capítulo 5, que contiene problemas concretos en espacios de dimensión infinita, para los cuales se puede aplicar la teoría antes descrita. Algunas referencias útiles sobre optimización convexa y no lineal son [2, 3, 7, 12].

En la Parte II estudiaremos algunos sistemas dinámicos que pueden utilizarse para encontrar soluciones para el problema de minimización convexa en espacios de Hilbert. El Capítulo 6 contiene una breve introducción al estudio asintótico de sistemas dinámicos clásicos, a tiempo continuo y discreto, que se pueden aplicar para resolver el problema de minimización convexa. Dichos sistemas se estudian detalladamente en los Capítulos 7, 8 y 9. Para terminar esta parte, en el Capítulo 10 presentamos algunas relaciones entre los sistemas estudiados antes.

Para comodidad del lector, incluimos además una breve discusión sobre las principales propiedades de los espacios de Hilbert en el Apéndice A.

Parte I

El Problema de Minimización

Capítulo 2

Conceptos generales

2.1 Funciones en los Reales Extendidos

El conjunto de los reales extendidos es una ampliación de \mathbf{R} , agregando el símbolo $+\infty$, y extendiendo el orden de manera que $\gamma < +\infty$ para todo $\gamma \in \mathbf{R}$. Esto nos permite definir funciones con valores en $\mathbf{R} \cup \{+\infty\}$. El ejemplo más sencillo es la llamada *función indicatriz* de un subconjunto C de otro conjunto X , definida como

$$\delta_C(x) = \begin{cases} 0 & \text{si } x \in C \\ +\infty & \text{si no.} \end{cases}$$

Claramente, si $f : X \rightarrow \mathbf{R}$, los problemas

$$\min\{f(x) : x \in C\} \quad \text{y} \quad \min\{f(x) + \delta_C(x) : x \in X\}$$

son equivalentes. La ventaja del segundo es que nos permite aprovechar diversos aspectos estructurales (que pueden ser geométricos, topológicos o de otra índole) de X .

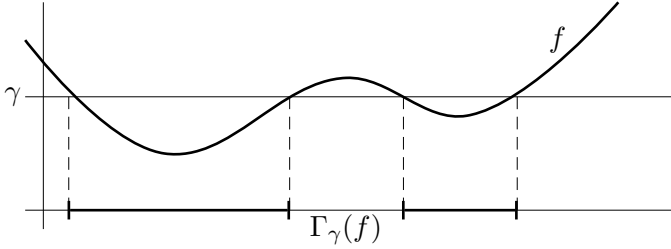
Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función. El *dominio efectivo* o, simplemente, *dominio* de f es el conjunto

$$\text{dom}(f) = \{x \in X : f(x) < +\infty\}.$$

Diremos que f es *propia* si $\text{dom}(f) \neq \emptyset$. En caso contrario, $f \equiv +\infty$, lo que no es muy interesante.

Dado $\gamma \in \mathbf{R}$ definimos el *conjunto de subnivel* γ de f como

$$\Gamma_\gamma(f) = \{x \in X : f(x) \leq \gamma\}.$$



Si $x \in \text{dom}(f)$, claramente $x \in \Gamma_{f(x)}(f)$. Por lo tanto,

$$\text{dom}(f) = \bigcup_{\gamma \in \mathbf{R}} \Gamma_\gamma(f).$$

Por otra parte, recordemos que

$$\text{argmin}(f) = \{x^* \in X : f(x^*) \leq f(x) \text{ para todo } x \in X\}.$$

Dado que $x \in \text{argmin}(f)$ si, y solo si, $x \in \Gamma_\gamma(f)$ para todo $\gamma > \inf(f)$, tenemos que

$$\text{argmin}(f) = \bigcap_{\gamma > \inf(f)} \Gamma_\gamma(f). \quad (2.1)$$

Finalmente, el *epígrafo* de f es el subconjunto del producto cartesiano $X \times \mathbf{R}$ definido por

$$\text{epi}(f) = \{(x, \alpha) \in X \times \mathbf{R} : f(x) \leq \alpha\}.$$

Este conjunto incluye al gráfico de f y a todos los puntos que están “por encima” de dicho gráfico.

2.2 Convexidad

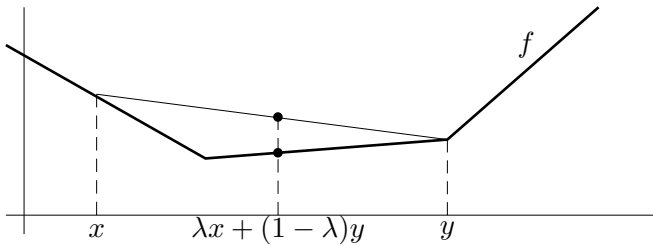
Sea X un espacio vectorial. Un conjunto $C \subseteq X$ es *convexo* si para cada par de puntos de C , el segmento que los une también pertenece a C .

En otras palabras, si el punto $\lambda x + (1 - \lambda)y$ pertenece a C siempre que $x, y \in C$ y $\lambda \in (0, 1)$.

Una función $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ es *convexa* si para cada $x, y \in X$ y cada $\lambda \in (0, 1)$ se cumple que

$$f\left(\lambda x + (1 - \lambda)y\right) \leq \lambda f(x) + (1 - \lambda)f(y), \quad (2.2)$$

o, de manera equivalente, si su epígrafo $\text{epi}(f)$ es un subconjunto convexo de $(X \times \mathbf{R})$.



Por otra parte, una función $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ es *cuasiconvexa* si para cada $\gamma \in \mathbf{R}$, el conjunto de subnivel $\Gamma_\gamma(f)$ es convexo.

Proposición 2.1. *Toda función convexa es cuasiconvexa.*

Demostración. Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa y sea $\gamma \in \mathbf{R}$. Debemos probar que para cada $x, y \in \Gamma_\gamma(f)$ y cada $\lambda \in (0, 1)$ se cumple que $\lambda x + (1 - \lambda)y \in \Gamma_\gamma(f)$. En efecto,

$$f\left(\lambda x + (1 - \lambda)y\right) \leq \lambda f(x) + (1 - \lambda)f(y) \leq \lambda \gamma + (1 - \lambda)\gamma = \gamma,$$

de donde se obtiene el resultado. □

El recíproco, sin embargo, es falso.

Ejemplo 2.2. Entre muchas otras, las funciones $f_1, \dots, f_4 : \mathbf{R} \rightarrow \mathbf{R}$, definidas, respectivamente, por:

1. $f_1(x) = x^3$,

2. $f_2(x) = \sqrt{|x|}$,
3. $f_3(x) = \exp(-\frac{1}{x^2})$, y
4. $f_4(x) = \lfloor x \rfloor = \text{máx}\{k \in \mathbf{Z} : k \leq x\}$,

son cuasiconvexas pero no convexas.

Proposición 2.3. *Si $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ es cuasiconvexa, entonces $\text{dom}(f)$ y $\text{argmin}(f)$ son conjuntos convexos.*

Demostración. Sean $x, y \in \text{dom}(f)$ y sea $\lambda \in (0, 1)$. Tenemos que

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

y así $\lambda x + (1 - \lambda)y \in \text{dom}(f)$.

De manera similar, si $x, y \in \text{argmin}(f)$ y $\lambda \in (0, 1)$, entonces

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) = \text{mín}(f).$$

Esto prueba que $\lambda x + (1 - \lambda)y \in \text{argmin}(f)$. □

Diremos que la función $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ es *estrictamente convexa* si la desigualdad (2.2) es estricta siempre que $x \neq y$, con $x, y \in \text{dom}(f)$.

Proposición 2.4. *Una función propia y estrictamente convexa tendrá, a lo más, un minimizador.*

Demostración. Supongamos que $x, y \in \text{argmin}(f)$. Si $x \neq y$, entonces

$$f\left(\frac{x + y}{2}\right) < \frac{f(x) + f(y)}{2} = \text{mín}\{f(z) : z \in X\},$$

lo cual es imposible. □

2.3 Minimización Local y Global

Si (X, τ) es un espacio vectorial topológico, la noción de vecindad nos permite distinguir entre minimizadores locales y globales. En efecto, sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función propia. Un punto $x^* \in \text{dom}(f)$ es un

minimizador global de f si $f(x^*) \leq f(x)$ para todo $x \in X$.¹ Por otra parte, diremos que x^* es un *minimizador local* de f si existe una vecindad \mathcal{V} de x^* tal que $f(x^*) \leq f(x)$ para todo $x \in \mathcal{V}$.

Claramente, todo minimizador global es también un minimizador local. El recíproco, en general, no es cierto.

Ejemplo 2.5. Todo número entero es minimizador local de la función $f : \mathbf{R} \rightarrow \mathbf{R}$ definida por $f(x) = 2\pi x + 2 \operatorname{sen}(2\pi x - \frac{5\pi}{6})$, pero ninguno de ellos es minimizador global pues $\lim_{x \rightarrow -\infty} f(x) = -\infty$.

Sin embargo, para las funciones convexas no hay distinción entre minimizadores locales y globales.

Proposición 2.6. *Sea (X, τ) un espacio vectorial topológico y sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa y propia. Si x^* es un minimizador local de f , entonces también es un minimizador global.*

Demostración. Sea \mathcal{V} una vecindad de x^* tal que $f(x^*) \leq f(x)$ para todo $x \in \mathcal{V}$, y supongamos que x^* no es un mínimo global. Es decir, existe $y \in X$ tal que $f(y) < f(x^*)$. Como (X, τ) es un espacio vectorial topológico, existe $\lambda \in (0, 1)$ tal que $x_\lambda = \lambda y + (1 - \lambda)x^* \in \mathcal{V}$. Entonces

$$f(x_\lambda) \leq \lambda f(y) + (1 - \lambda)f(x^*) < \lambda f(x^*) + (1 - \lambda)f(x^*) = f(x^*),$$

lo que es una contradicción. □

¹O, de manera equivalente, para todo $x \in \operatorname{dom}(f)$.

Capítulo 3

Existencia de Soluciones

3.1 El Teorema de Weierstrass-Hilbert-Tonelli

El Teorema de Weierstrass (Teorema 1.1) se apoya en dos hipótesis sobre el problema: la continuidad de la función objetivo y la compacidad del dominio. En esta sección estudiaremos una versión más general de ese resultado.

A lo largo de esta sección, (X, τ) es un espacio de Hausdorff¹. Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función y sea $x_0 \in X$. Diremos que f es *inferiormente semicontinua* en x_0 si para cada $\alpha < f(x_0)$ existe una vecindad \mathcal{V} de x_0 tal que $f(y) > \alpha$ para todo $y \in \mathcal{V}$. Si f es inferiormente semicontinua en cada punto de X , diremos simplemente que f es inferiormente semicontinua.

Proposición 3.1. *Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$. Las siguientes afirmaciones son equivalentes:*

- i) La función f es inferiormente semicontinua;*
- ii) El conjunto $\text{epi}(f)$ es cerrado en $X \times \mathbf{R}$; y*
- iii) Para cada $\gamma \in \mathbf{R}$, el conjunto de subnivel $\Gamma_\gamma(f)$ es cerrado.*

Demostración. Probaremos que i) \Rightarrow ii) \Rightarrow iii) \Rightarrow i).

Suponamos que f es inferiormente semicontinua y sea $(x_0, \alpha) \notin \text{epi}(f)$.

¹Un espacio topológico en el que puntos distintos admiten vecindades disjuntas.

Entonces $\alpha < f(x_0)$. Tomemos cualquier $\beta \in (\alpha, f(x_0))$. Sabemos que existe una vecindad \mathcal{V} de x_0 tal que $f(y) > \beta$ para todo $y \in \mathcal{V}$. El conjunto $\mathcal{V} \times (-\infty, \beta)$ es una vecindad de (x_0, α) que no interseca a $\text{epi}(f)$, lo que prueba ii).

Supongamos ahora que $\text{epi}(f)$ es cerrado. Es fácil ver que para cada $\gamma \in \mathbf{R}$, el conjunto $\Gamma_\gamma(f)$ es homeomorfo a $\text{epi}(f) \cap [X \times \{\gamma\}]$ y, por lo tanto, es cerrado.

Finalmente, supongamos que $\Gamma_\gamma(f)$ es cerrado y probemos que f es inferiormente semicontinua. Para ello, escojamos $x_0 \in X$ y $\alpha \in \mathbf{R}$ tales que $\alpha < f(x_0)$. Entonces $x_0 \notin \Gamma_\alpha(f)$. Como este conjunto es cerrado, existe una vecindad \mathcal{V} de x_0 que no interseca a $\Gamma_\alpha(f)$. En otras palabras, $f(y) > \alpha$ para todo $y \in \mathcal{V}$. \square

Una función $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ es *inf-compacta* si para cada $\gamma \in \mathbf{R}$ el conjunto de subnivel $\Gamma_\gamma(f)$ es relativamente compacto.

Corolario 3.2. *Una función propia $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ es inferiormente semicontinua e inf-compacta si, y solo si, $\Gamma_\gamma(f)$ es compacto y no vacío para cada $\gamma > \inf(f)$.*

Teorema 3.3 (Teorema de Weierstrass-Hilbert-Tonelli). *Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función propia, inferiormente semicontinua e inf-compacta. Entonces el conjunto $\text{argmin}(f)$ es compacto y no vacío.*

Demostración. En virtud de (2.1), basta observar que cada $\Gamma_\gamma(f)$ es compacto y no vacío, y recordar que la intersección decreciente de compactos no vacíos da un conjunto compacto y no vacío. \square

Inmediatamente obtenemos el Teorema 1.1:

Corolario 3.4 (Teorema de Weierstrass). *Si C es un subconjunto cerrado, acotado y no vacío de \mathbf{R}^N y $f : C \rightarrow \mathbf{R}$ es una función continua, entonces f alcanza su máximo y su mínimo en C .*

Demostración. La función $g : \mathbf{R}^N \rightarrow \mathbf{R} \cup \{+\infty\}$, definida por

$$g(x) = \begin{cases} f(x) & \text{si } x \in C \\ +\infty & \text{si no,} \end{cases}$$

es propia, inferiormente semicontinua e inf-compacta para la topología definida por la norma en \mathbf{R}^N . El Teorema 3.3 garantiza la existencia de

un minimizador para g en \mathbf{R}^N , el cual es también un minimizador para f en C . Para el maximizador, basta aplicar el mismo argumento con $-f$ en lugar de f . \square

Entre las hipótesis de Teorema 3.3 se produce una tensión: mientras más conjuntos abiertos tenga la topología τ , habrá más funciones inferiormente semicontinuas, pero menos subconjuntos compactos y, por lo tanto, menos funciones inf-compactas. Más adelante veremos que, para el caso convexo, hay una topología que hace más sencilla la verificación de las hipótesis del Teorema 3.3 (ver Teorema 3.13).

3.2 Breve Repaso de Análisis Funcional

En esta sección recordaremos, sin entrar en demasiados detalles, algunos conceptos y resultados de Análisis Funcional que utilizaremos más adelante. En particular, omitiremos las demostraciones de los Teoremas 3.5 y 3.9 pues van más allá de los objetivos de este curso. Una presentación más exhaustiva se puede encontrar, por ejemplo, en [5].

Sea $(X, \|\cdot\|)$ un espacio vectorial normado. En lo que sigue llamaremos *topología fuerte* a la topología definida en X por la norma $\|\cdot\|$.

El *dual topológico* de X es el espacio vectorial X^* de las funciones, también llamadas *funcionales*, que son lineales y continuas de X (con la topología fuerte) en \mathbf{R} (con la topología usual). Denotemos por $\langle L, x \rangle$ la acción de $L \in X^*$ sobre $x \in X$. En X^* se puede definir la norma $\|L\|_* = \sup_{\|x\|=1} \langle L, x \rangle$, de manera que $\langle L, x \rangle \leq \|L\|_* \|x\|$ para cada $x \in X$ y $L \in X^*$. Con esto, $(X^*, \|\cdot\|_*)$ es un espacio vectorial normado. Más aun, siempre es un espacio de Banach².

Por definición, cada elemento de X^* es continuo como función de X (con la topología fuerte) en \mathbf{R} (con la topología usual). Sin embargo, podrían existir otras topologías en X que también hagan continuos a todos los elementos de X^* . A la más gruesa (aquella que tenga menos abiertos) entre todas ellas la llamaremos *topología débil*. Dados $x_0 \in X$, $L \in X^*$

²Aunque $(X, \|\cdot\|)$ no sea completo.

y $\varepsilon > 0$, cada conjunto de la forma

$$\mathcal{V}_L^\varepsilon(x_0) = \{x \in X : \langle L, x - x_0 \rangle < \varepsilon\}$$

es abierto para la topología débil y contiene a x_0 . Más aun, ellos generan una base de vecindades para x_0 en el siguiente sentido: Si V_0 es una vecindad de x_0 , entonces existen $L_1, \dots, L_N \in X^*$ y $\varepsilon > 0$ tales que

$$x_0 \in \bigcap_{k=1}^N \mathcal{V}_{L_k}^\varepsilon(x_0) \subset V_0.$$

Se puede demostrar que si la dimensión de X es finita, la topología débil coincide con la fuerte. Sin embargo, si la dimensión es infinita, la topología débil es estrictamente más gruesa que la fuerte. En otras palabras, todo conjunto que sea abierto para la topología fuerte es también abierto para la topología débil, pero existen conjuntos que son abiertos para la topología fuerte, y que no lo son para la topología débil. Un ejemplo es la bola unitaria $B(0, 1)$. Por supuesto, lo mismo vale para los conjuntos cerrados. En la Proposición 3.7 veremos que para los conjuntos convexos, en cambio, no hay distinción entre ser cerrados para la topología fuerte o la débil. Comenzaremos por enunciar el siguiente resultado clásico de Análisis Funcional:

Teorema 3.5 (Teorema de Hahn-Banach). *Sean A y B subconjuntos convexos, no vacíos y disjuntos de un espacio vectorial normado X .*

- i) Si A es abierto, entonces existe $L \in X^* \setminus \{0\}$ tal que $\langle L, x \rangle \leq \langle L, y \rangle$ para cada $x \in A$ e $y \in B$.*
- ii) Si A es compacto y B es cerrado, entonces existen $L \in X^* \setminus \{0\}$ y $\varepsilon > 0$ tales que $\langle L, x \rangle + \varepsilon \leq \langle L, y \rangle$ para cada $x \in A$ e $y \in B$.*

Veremos dos consecuencias importantes:

Proposición 3.6. *Sea X un espacio vectorial normado. La topología débil es de Hausdorff.*

Demostración. Sean $x_1 \neq x_2$. La parte ii) del Teorema 3.5 asegura la existencia de $L \in X^*$ tal que $\langle L, x_1 \rangle + \varepsilon \leq \langle L, x_2 \rangle$. Entonces $\mathcal{V}_L^{\frac{\varepsilon}{2}}(x_1)$ es vecindad de x_1 , $\mathcal{V}_{-L}^{\frac{\varepsilon}{2}}(x_2)$ es vecindad de x_2 y $\mathcal{V}_L^{\frac{\varepsilon}{2}}(x_1) \cap \mathcal{V}_{-L}^{\frac{\varepsilon}{2}}(x_2) = \emptyset$. \square

Proposición 3.7. *Sea C un subconjunto convexo de un espacio vectorial normado X . Entonces C es cerrado para la topología fuerte si, y solo si, es cerrado para la topología débil.*

Demostración. Como la topología débil es más gruesa que la fuerte, todo cerrado débil es también cerrado fuerte. Para el recíproco, supongamos que C es convexo y cerrado fuerte. Dado $x_0 \notin C$, aplicamos la parte ii) del Teorema 3.5 con $A = \{x_0\}$ y $B = C$ para deducir que existen $L \in X^* \setminus \{0\}$ y $\varepsilon > 0$ tales que $\langle L, x_0 \rangle + \varepsilon \leq \langle L, y \rangle$ para todo $y \in C$. El conjunto $\mathcal{V} = \{x \in X : \langle L, x - x_0 \rangle < \varepsilon\}$ es una vecindad de x_0 para la topología débil y no intersecta a C . Como esto se puede hacer para cualquier $x_0 \notin C$, concluimos que C es cerrado débil. \square

Una ventaja de las funciones cuasiconvexas respecto de la semicontinuidad inferior viene dada por la siguiente:

Proposición 3.8. *Sea $(X, \|\cdot\|)$ un espacio vectorial normado y sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función cuasiconvexa. Entonces f es inferiormente semicontinua para la topología fuerte si, y solo si, es inferiormente semicontinua para la topología débil.*

Demostración. Es inmediato de las Proposiciones 3.1 y 3.7. \square

En particular, si una función cuasiconvexa es continua para la topología fuerte, ella es inferiormente semicontinua para la topología débil.

En general, la compacidad es una propiedad difícil de verificar. Sin embargo, el siguiente resultado nos da una caracterización útil, cuando el espacio es reflexivo. Una sucesión $(x_n)_{n \in \mathbf{N}}$ en X converge débilmente a \bar{x} si $\lim_{n \rightarrow \infty} \langle L, x_n - \bar{x} \rangle = 0$ para cada $L \in X^*$. En cambio, si $\lim_{n \rightarrow \infty} \|x_n - \bar{x}\| = 0$ diremos que la sucesión converge fuertemente a \bar{x} .

Teorema 3.9 (Teorema de Kakutani-Eberlein-Šmulian). *Sea $(X, \|\cdot\|)$ un espacio de Banach. Las siguientes afirmaciones son equivalentes:*

- i) *El espacio X es reflexivo.*
- ii) *La bola unitaria cerrada $\bar{B}(0, 1)$ es compacta para la topología débil. En particular, todo conjunto acotado es relativamente compacto para la topología débil.*

iii) Toda sucesión acotada en X tiene alguna subsucesión débilmente convergente.

Corolario 3.10. Sea $(X, \|\cdot\|)$ un espacio de Banach reflexivo y sea $y : \mathbf{R}_+ \rightarrow X$ una función acotada. Entonces $y(t)$ converge débilmente si, y sólo si, tiene un único punto de acumulación débil, cuando $t \rightarrow \infty$.

Demostración. Si $y(t)$ converge débilmente a \bar{y} cuando $t \rightarrow \infty$, entonces \bar{y} es su único punto de acumulación débil. Para el recíproco, sea \hat{y} el único punto de acumulación débil de $y(t)$ y supongamos que $y(t)$ no converge débilmente a \hat{y} cuando $t \rightarrow \infty$. Entonces existe una vecindad débil \mathcal{V} de \hat{y} y una sucesión $(t_n)_{n \in \mathbf{N}}$ tal que $y(t_n) \notin \mathcal{V}$ para todo $n \in \mathbf{N}$. Como $(y(t_n))_{n \in \mathbf{N}}$ es acotada, tiene una subsucesión que converge débilmente cuando $n \rightarrow \infty$ a un $\tilde{y} \neq \hat{y}$. Esto contradice la unicidad de \hat{y} . \square

Diremos que f es *coerciva* si para cada $\gamma \in \mathbf{R}$ el conjunto de subnivel $\Gamma_\gamma(f)$ es acotado. Dado que los conjuntos compactos son acotados, toda función inf-compacta es coerciva. El recíproco no es cierto en general, pero tenemos el siguiente resultado, consecuencia del Teorema 3.9:

Corolario 3.11. En un espacio de Banach reflexivo, toda función coerciva es inf-compacta para la topología débil.

3.3 Minimización Convexa

Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función propia definida sobre cualquier conjunto X , y sea (x_n) una sucesión en $\text{dom}(f)$. Diremos que (x_n) es una *sucesión minimizante* para f si $\lim_{n \rightarrow \infty} f(x_n) = \inf(f)$. De acuerdo con la definición de ínfimo, siempre existe alguna sucesión minimizante, entendiéndose que el $\lim_{n \rightarrow \infty} f(x_n)$ puede ser $-\infty$. Claramente, si (x_n) es una sucesión minimizante para f , entonces existe $\gamma \in \mathbf{R}$ tal que $x_n \in \Gamma_\gamma(f)$ para todo $n \in \mathbf{N}$.

Proposición 3.12. Sea $(X, \|\cdot\|)$ un espacio vectorial normado y sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función cuasiconvexa, propia e inferiormente semicontinua. Sea (x_n) una sucesión minimizante para f .

i) Si x_n converge débilmente a \bar{x} , entonces $\bar{x} \in \text{argmin}(f)$.

ii) Si X es un espacio de Banach reflexivo y (x_n) es acotada, entonces $\text{argmin}(f) \neq \emptyset$.

Demostración. Como f es cuasiconvexa e inferiormente semicontinua, f también es inferiormente semicontinua para la topología débil, en virtud de la Proposición 3.8. Para probar i), supongamos que x_n converge débilmente a \bar{x} . Entonces

$$f(\bar{x}) \leq \liminf_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} f(x_n) = \inf(f)$$

y $\bar{x} \in \operatorname{argmin}(f)$. Para la parte ii), si (x_n) es acotada, El Teorema 3.9 nos dice que ella tiene una subsucesión (x_{k_n}) que converge débilmente a algún \hat{x} . Como dicha subsucesión también es minimizante, la parte i) nos dice que $\hat{x} \in \operatorname{argmin}(f)$. \square

Estamos ahora en posición de demostrar nuestro principal resultado de existencia de soluciones para el problema de minimización de una función cuasiconvexa:

Teorema 3.13. *Sea $(X, \|\cdot\|)$ un espacio de Banach reflexivo y sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función cuasiconvexa, propia, inferiormente semicontinua y coerciva. Entonces $\operatorname{argmin}(f)$ es débilmente compacto, no vacío y convexo.*

Presentaremos dos demostraciones: la primera se basa en el Teorema 3.3 y la segunda en el llamado *método directo*, que consiste en encontrar minimizadores como puntos de acumulación de sucesiones minimizantes.

Primera Demostración. Como f es cuasiconvexa e inferiormente semicontinua, para cada $\gamma \in \mathbf{R}$ el conjunto de subnivel $\Gamma_\gamma(f)$ es convexo y cerrado. Luego también es cerrado para la topología débil. Por otra parte, como f es coerciva, $\Gamma_\gamma(f)$ es acotado. Dado que el espacio es reflexivo, $\Gamma_\gamma(f)$ es compacto para la topología débil. Basta usar el Teorema 3.3 en el espacio topológico (X, τ) , donde τ es la topología débil. La Proposición 2.3 establece que $\operatorname{argmin}(f)$ es convexo. \square

Segunda Demostración. Sea (x_n) una sucesión minimizante. Como f es coerciva, (x_n) es acotada. La parte ii) de la Proposición 3.12 nos dice que $\operatorname{argmin}(f) \neq \emptyset$. Se concluye como en la primera demostración. \square

De acuerdo con la Proposición 3.8 y el Teorema 3.13, es importante tener condiciones que garanticen la semicontinuidad inferior de las funciones convexas y cuasiconvexas, al menos para la topología fuerte. Daremos una respuesta en la Proposición 4.15.

Finalmente, presentamos un resultado que establece existencia y unicidad de solución para el problema de minimización:

Corolario 3.14. *Sea X un espacio de Banach reflexivo y sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función estrictamente convexa, inferiormente semicontinua, propia y coerciva. Entonces f tiene exactamente un minimizador.*

Demostración. La existencia se obtiene de aplicar la Proposición 2.1 y el Teorema 3.13. La unicidad se deduce de la Proposición 2.4. \square

Capítulo 4

Optimalidad

En todo este capítulo, $(X, \|\cdot\|)$ un espacio vectorial normado sobre \mathbf{R} . Dados $x_0 \in X$ y $r > 0$, denotaremos por $B(x_0, r)$ y $\bar{B}(x_0, r)$, respectivamente, las bolas abierta y cerrada con centro x_0 y radio r .

4.1 Caso Diferenciable

Sea A un subconjunto no vacío de X . Una función $f : A \rightarrow \mathbf{R}$ es *continua* en $a \in A$ si para cada $\varepsilon > 0$ existe $\delta > 0$ tal que $|f(x) - f(a)| < \varepsilon$ para todo $x \in A$ tal que $\|x - a\| < \delta$. Si f es continua en todo punto de A decimos simplemente que f es *continua en A* . Diremos además que f es *Lipschitz-continua* con constante $M > 0$ si

$$|f(x) - f(y)| \leq M\|x - y\|$$

para todo $x, y \in A$. Toda función Lipschitz-continua es continua.

Supongamos ahora que A es abierto. Una función $f : A \rightarrow \mathbf{R}$ es *diferenciable en el sentido de Fréchet* (o simplemente *diferenciable*) en un punto $a \in A$ si existe un funcional lineal continuo¹ $L \in X^*$ tal que

$$\lim_{\|h\| \rightarrow 0} \frac{|f(a+h) - f(a) - \langle L, h \rangle|}{\|h\|} = 0.$$

¹Recordemos la notación de la Sección 3.2.

En ese caso, decimos que L es la *derivada de Fréchet* de f en a y la denotamos por $Df(a)$. Si f es diferenciable en todo punto a de A decimos que f es *diferenciable en A* . Una consecuencia inmediata de la diferenciabilidad es la siguiente:

Proposición 4.1. *Si f es diferenciable en a , también es continua en a .*

Demostración. Simplemente notemos que si $Df(a) = L$ entonces

$$|f(a+h) - f(a)| \leq \frac{|f(a+h) - f(a) - \langle L, h \rangle|}{\|h\|} \|h\| + \|L\|_* \|h\|.$$

El lado derecho tiende a 0 cuando $\|h\|$ tiende a 0. □

Esta definición de diferenciabilidad nos permite extender el Teorema 1.2 a cualquier espacio vectorial normado:

Teorema 4.2 (Regla de Fermat). *Sea $(X, \|\cdot\|)$ un espacio vectorial normado y sea A un subconjunto abierto de X . Si $x^* \in A$ es un minimizador local de $f : A \rightarrow \mathbf{R}$ y si f es diferenciable en x^* , entonces $Df(x^*) = 0$.²*

Demostración. Sea $r > 0$ tal que $B(x^*, r) \subset A$ y $f(x^*) \leq f(x)$ para todo $x \in B(x^*, r)$. Sea $h \in B(0, 1)$. Dado que $Df(x^*)$ es lineal, para cada $\rho \in (0, r)$ se cumple que

$$\langle Df(x^*), h \rangle = \frac{1}{\rho} \langle Df(x^*), \rho h \rangle \geq \frac{\langle Df(x^*), \rho h \rangle + f(x^*) - f(x^* + \rho h)}{\rho}.$$

Tomando el límite cuando ρ tiende a 0 vemos que $\langle Df(x^*), h \rangle \geq 0$ para todo $h \in B(0, 1)$. Cambiando h por $-h$ se demuestra que $\langle Df(x^*), h \rangle \leq 0$ para todo $h \in B(0, 1)$ y concluimos que $Df(x^*) = 0$. □

Observación 4.3. Si $X = \mathbf{R}^N$, la definición de diferenciabilidad dada arriba coincide con la definición usual que se estudia en los cursos de Cálculo en Varias Variables. Si $f : A \subseteq \mathbf{R}^N \rightarrow \mathbf{R}$ es diferenciable en $a \in A$ y

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), \dots, \frac{\partial f}{\partial x_N}(a) \right)$$

²Esto quiere decir que $\langle Df(x^*), h \rangle = 0$ para toda $h \in X$.

es el *gradiente* de f en a , entonces

$$\langle Df(a), h \rangle = \nabla f(a) \cdot h$$

para cada $h \in \mathbf{R}^N$.

Teniendo en cuenta lo anterior, el Teorema 1.2 es un caso particular del Teorema 4.2:

Corolario 4.4 (Regla de Fermat en \mathbf{R}^N). *Sea A un subconjunto abierto de \mathbf{R}^N . Si $x^* \in A$ es un minimizador de $f : A \rightarrow \mathbf{R}$ en A y f es diferenciable en x^* , entonces $\nabla f(x^*) = 0$.*

Demostración. En virtud del Teorema 4.2 y la Observación 4.3, tenemos que $\nabla f(x^*) \cdot h = 0$ para todo $h \in \mathbf{R}^N$ y, por lo tanto, $\nabla f(x^*) = 0$. \square

4.2 Caso Convexo, no Diferenciable

Los resultados de la sección anterior pueden aplicarse cuando el minimizador local se encuentra en el interior del conjunto factible y la función objetivo es diferenciable en dicho punto. Sin embargo, este no siempre es el caso. Veremos ahora una generalización del concepto de derivada que es especialmente útil para caracterizar minimizadores de funciones convexas.

4.2.1 Subdiferencial y Regla de Fermat Generalizada

Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa y propia. El *subdiferencial* de f en x es el conjunto

$$\partial f(x) = \{x^* \in X^* : f(y) \geq f(x) + \langle x^*, y - x \rangle \text{ para todo } y \in X\}.$$

Los elementos de $\partial f(x)$ se llaman *subgradientes* de f en x . Si $\partial f(x) \neq \emptyset$ decimos que f es *subdiferenciable* en x . El *dominio* del subdiferencial es el conjunto

$$\text{dom}(\partial f) = \{x \in X : \partial f(x) \neq \emptyset\}.$$

Claramente, $\text{dom}(\partial f) \subset \text{dom}(f)$. Además, el subdiferencial extiende la noción de derivada en el siguiente sentido:

Proposición 4.5. *Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa y propia. Si f es diferenciable en x , entonces $x \in \text{dom}(\partial f)$ y $\partial f(x) = \{Df(x)\}$.*

Demostración. Sea $x^* \in \partial f(x)$. Como f es diferenciable en x , para cada $\varepsilon > 0$ existe $\delta > 0$ tal que

$$f(y) \leq f(x) + \langle Df(x), y - x \rangle + \varepsilon \|y - x\|$$

para cada $y \in B(x, \delta)$. Por otra parte, como $x^* \in \partial f(x)$, tenemos que

$$f(y) \geq f(x) + \langle x^*, y - x \rangle$$

para todo $y \in X$. Por lo tanto,

$$\langle x^* - Df(x), y - x \rangle \leq \varepsilon \|y - x\|$$

para cada $y \in B(x, \delta)$. Esto implica que

$$\langle x^* - Df(x), h \rangle \leq \varepsilon$$

para cada $h \in B(0, 1)$. Por lo tanto, $\|x^* - Df(x)\|_* \leq \varepsilon$. Como esto es válido para cada $\varepsilon > 0$ concluimos que $x^* = Df(x)$. \square

Observación 4.6. Para fijar ideas, los subgradientes de una función $f : \mathbf{R} \rightarrow \mathbf{R}$ son las pendientes de las rectas que tocan al gráfico de f en el punto $(x, f(x))$ y que están siempre por debajo de dicho gráfico.

De la definición se desprende inmediatamente el siguiente resultado:

Teorema 4.7 (Regla de Fermat Generalizada). *Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa. Entonces \hat{x} es un minimizador global de f si, y sólo si, $0 \in \partial f(\hat{x})$.*

Este resultado generaliza la Regla de Fermat usual a funciones convexas, pero no diferenciables, en cualquier espacio vectorial normado

Observación 4.8. Si además f es diferenciable en \hat{x} , teniendo en cuenta la Proposición 4.5, vemos que \hat{x} es un minimizador global de f si, y sólo si, $0 = Df(\hat{x})$. Esto nos dice que para funciones convexas, la condición necesaria de optimalidad que establece la Regla de Fermat es también suficiente.

Veamos algunos ejemplos sencillos:

Ejemplo 4.9. Para la función $f : \mathbf{R} \rightarrow \mathbf{R}$ dada por $f(x) = |x|$, tenemos

$$\partial f(x) = \begin{cases} \{-1\} & \text{si } x < 0 \\ [-1, 1] & \text{si } x = 0 \\ \{1\} & \text{si } x > 0. \end{cases}$$

Ejemplo 4.10. Sea $g : \mathbf{R} \rightarrow \mathbf{R} \cup \{+\infty\}$, la función definida mediante $g(x) = +\infty$ si $x < 0$ y $g(x) = -\sqrt{x}$, si $x \geq 0$. Entonces

$$\partial g(x) = \begin{cases} \emptyset & \text{si } x \leq 0 \\ \left\{ -\frac{1}{2\sqrt{x}} \right\} & \text{si } x > 0. \end{cases}$$

Notemos que $0 \in \text{dom}(g)$ pero $\partial g(0) = \emptyset$. Vemos aquí un caso donde $\text{dom}(\partial g) \subsetneq \text{dom}(g)$.

Ejemplo 4.11. Si definimos $h : \mathbf{R} \rightarrow \mathbf{R} \cup \{+\infty\}$ como $h(x) = +\infty$ si $x \neq 0$ y $h(0) = 0$, entonces

$$\partial h(x) = \begin{cases} \emptyset & \text{si } x \neq 0 \\ \mathbf{R} & \text{si } x = 0. \end{cases}$$

Ejemplo 4.12. Más generalmente, sea C un subconjunto convexo, cerrado y no vacío de X y sea $\delta_C : X \rightarrow \mathbf{R} \cup \{+\infty\}$ la función indicatriz del conjunto C (ver Sección 2.1):

$$\delta_C(x) = \begin{cases} 0 & \text{si } x \in C \\ +\infty & \text{si no.} \end{cases}$$

El conjunto $N_C(x) = \partial \delta_C(x)$ es el *cono normal* a C en x . Está dado por

$$N_C(x) = \{x^* \in X^* : \langle x^*, y - x \rangle \leq 0 \text{ para todo } y \in C\}$$

si $x \in C$, y $N_C(x) = \emptyset$ si $x \notin C$. Intuitivamente, el cono normal contiene las direcciones que apuntan “hacia afuera” de C .

Proposición 4.13. Para cada $x \in X$, el conjunto $\partial f(x)$ es convexo y cerrado (posiblemente vacío).

Demostración. Sean $x_1^*, x_2^* \in \partial f(x)$ y $\lambda \in (0, 1)$. Para cada $y \in X$ tenemos que

$$\begin{aligned} f(y) &\geq f(x) + \langle x_1^*, y - x \rangle \\ f(y) &\geq f(x) + \langle x_2^*, y - x \rangle \end{aligned}$$

Combinando adecuadamente estas dos desigualdades obtenemos

$$f(y) \geq f(x) + \langle \lambda x_1^* + (1 - \lambda)x_2^*, y - x \rangle$$

para cada $y \in X$, lo que nos dice que $\lambda x_1^* + (1 - \lambda)x_2^* \in \partial f(x)$.

Para ver que $\partial f(x)$ es cerrado, tomemos una sucesión (x_n^*) de elementos de $\partial f(x)$, que converge a algún \hat{x}^* . Tenemos que

$$f(y) \geq f(x) + \langle x_n^*, y - x \rangle$$

para cada $y \in X$ y cada $n \in \mathbf{N}$. Cuando $n \rightarrow \infty$ obtenemos

$$f(y) \geq f(x) + \langle \hat{x}^*, y - x \rangle.$$

Concluimos que $\hat{x}^* \in \partial f(x)$. □

Recordemos que si $f : \mathbf{R} \rightarrow \mathbf{R}$ es una función convexa y diferenciable, entonces su derivada es una función creciente. El siguiente resultado generaliza esta propiedad.

Proposición 4.14. *Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa y sean $x, y \in \text{dom}(\partial f)$. Entonces*

$$\langle x^* - y^*, x - y \rangle \geq 0$$

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

Demostración. Si $x^* \in \partial f(x)$ y $y^* \in \partial f(y)$, de acuerdo con la definición del subdiferencial, tenemos que

$$\begin{aligned} f(y) &\geq f(x) + \langle x^*, y - x \rangle \\ f(x) &\geq f(y) + \langle y^*, x - y \rangle. \end{aligned}$$

Sumando estas dos desigualdades y agrupando los términos obtenemos $\langle x^* - y^*, x - y \rangle \geq 0$, como queríamos. □

4.2.2 Continuidad y Subdiferenciabilidad

Presentaremos ahora algunas propiedades de continuidad y subdiferenciabilidad, que revelan lo notablemente regulares que son las funciones convexas. Algunas demostraciones son un poco más sofisticadas, por lo que pueden omitirse en una primera lectura.

Proposición 4.15. *Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa y sea $x_0 \in X$. Las siguientes afirmaciones son equivalentes:*

- i) f está acotada superiormente en alguna vecindad de x_0 ;
- ii) f es Lipschitz-continua en alguna vecindad de x_0 ;
- iii) f es continua en x_0 ; y
- iv) $(x_0, \lambda) \in \text{int}(\text{epi}(f))$ para cada $\lambda > f(x_0)$.

Demostración. Probaremos que i) \Rightarrow ii) \Rightarrow iii) \Rightarrow iv) \Rightarrow i).

i) \Rightarrow ii): Si f está acotada superiormente en una vecindad de x_0 , existen $r > 0$ y $K > f(x_0)$ tales que $f(z) \leq K$ para todo $z \in B(x_0, 2r)$. Probaremos que existe $M > 0$ tal que $|f(x) - f(y)| \leq M\|x - y\|$ para cada $x, y \in B(x_0, r)$. En primer lugar, consideremos el punto

$$\tilde{y} = y + r \frac{y - x}{\|y - x\|}. \quad (4.1)$$

Por una parte, tenemos que $\|\tilde{y} - x_0\| \leq \|y - x_0\| + r < 2r$, por lo que $f(\tilde{y}) \leq K$. Por otra parte, despejando y en (4.1), vemos que

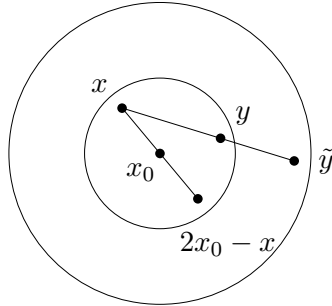
$$y = \lambda \tilde{y} + (1 - \lambda)x, \quad \text{con} \quad \lambda = \frac{\|y - x\|}{\|y - x\| + r} \leq \frac{\|y - x\|}{r}.$$

Por la convexidad de f ,

$$f(y) - f(x) \leq \lambda[f(\tilde{y}) - f(x)] \leq \lambda[K - f(x)]. \quad (4.2)$$

Como $x_0 = \frac{1}{2}x + \frac{1}{2}(2x_0 - x)$, tenemos que $f(x_0) \leq \frac{1}{2}f(x) + \frac{1}{2}f(2x_0 - x)$. Luego,

$$-f(x) \leq K - 2f(x_0). \quad (4.3)$$



Combinando (4.2) y (4.3), deducimos que

$$f(y) - f(x) \leq 2\lambda[K - f(x_0)] \leq \frac{2(K - f(x_0))}{r} \|x - y\|.$$

Intercambiando los roles de x e y vemos que $|f(x) - f(y)| \leq M\|x - y\|$ con $M = \frac{2(K - f(x_0))}{r} > 0$.

ii) \Rightarrow iii): Es inmediato.

iii) \Rightarrow iv): Si f es continua en x_0 y $\lambda > f(x_0)$, entonces existe $\delta > 0$ tal que $f(z) < \frac{\lambda}{2}$ para todo $z \in B(x_0, \delta)$. Por lo tanto,

$$(x_0, \lambda) \in B(x_0, \delta) \times \left(\frac{\lambda}{2}, +\infty\right) \subset \text{epi}(f),$$

con lo cual $(x_0, \lambda) \in \text{int}(\text{epi}(f))$.

iv) \Rightarrow i): Si $(x_0, \lambda) \in \text{int}(\text{epi}(f))$, entonces existen $r > 0$ y $K > f(x_0)$, tales que $B(x_0, r) \times (K, +\infty) \subset \text{epi}(f)$. Esto dice que $f(z) \leq K$ para todo $z \in B(x_0, r)$. \square

Observación 4.16. Se puede demostrar, usando la Proposición 4.15, que si $f : \mathbf{R}^N \rightarrow \mathbf{R} \cup \{+\infty\}$ es convexa, entonces ella es continua en $\text{int}(\text{dom}(f))$.

Ya hemos visto, en la Proposición 2.6 que algunas propiedades locales se vuelven globales cuando las funciones son convexas. Veremos ahora dos ejemplos más:

Corolario 4.17. Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa. Si f es continua en algún $x_0 \in \text{dom}(f)$, entonces $x_0 \in \text{int}(\text{dom}(f))$ y f es continua en todo $\text{int}(\text{dom}(f))$.

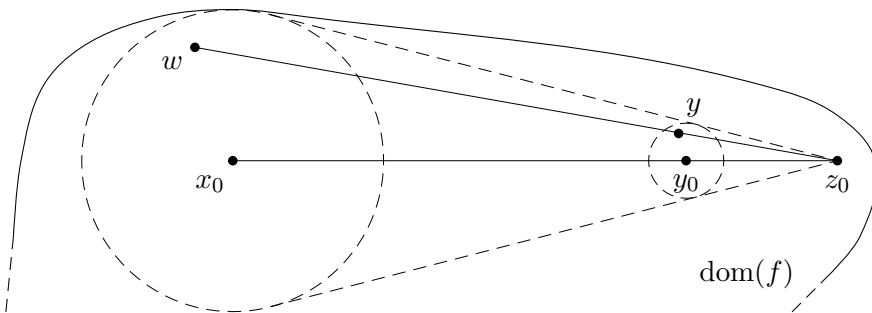
Demostración. Si f es continua en x_0 , la Proposición 4.15 nos dice que $x_0 \in \text{int}(\text{dom}(f))$ y que existen $r > 0$ y $K > f(x_0)$ tales que $f(x) \leq K$ para todo $x \in B(x_0, r)$. Sea $y_0 \in \text{int}(\text{dom}(f))$ y escojamos $\rho > 0$ de manera que el punto $z_0 = y_0 + \rho(y_0 - x_0)$ pertenezca a $\text{dom}(f)$. Tomemos $y \in B(y_0, \frac{\rho r}{1+\rho})$ y definamos $w = x_0 + (\frac{1+\rho}{\rho})(y - y_0)$. En otras palabras,

$$y = \left(\frac{\rho}{1+\rho}\right)w + \left(\frac{1}{1+\rho}\right)z_0.$$

Por otra parte,

$$\|w - x_0\| = \left(\frac{1+\rho}{\rho}\right)\|y - y_0\| < r,$$

con lo cual $w \in B(x_0, r)$ y $f(w) \leq K$.



Así,

$$f(y) \leq \left(\frac{\rho}{1+\rho}\right)f(w) + \left(\frac{1}{1+\rho}\right)f(z_0) \leq \max\{K, f(z_0)\}.$$

Como esto vale para cada $y \in B(y_0, \frac{\rho r}{1+\rho})$, concluimos que f está acotada superiormente en una vecindad de y_0 . La Proposición 4.15 nos garantiza que f es continua en y_0 . \square

Proposición 4.18. Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa. Entonces $x^* \in \partial f(x)$ si, y solo si, $f(y) \geq f(x) + \langle x^*, y - x \rangle$ para todo y en una vecindad de x .

Demostración. Si $x^* \in \partial f(x)$, claramente se tiene que $f(y) \geq f(x) + \langle x^*, y - x \rangle$ para todo y en una vecindad de x (todo el espacio). Para el recíproco, podemos suponer que existe $r > 0$ tal que $f(y) \geq f(x) + \langle x^*, y - x \rangle$ para todo $y \in B(x, r)$. Debemos probar que $x^* \in \partial f(x)$.

Sea $z \in X$ y tomemos $\lambda \in (0, 1)$ tal que $\lambda\|z - x\| < r$, de manera que $\lambda z + (1 - \lambda)x \in B(x, r)$. Así,

$$\lambda f(z) + (1 - \lambda)f(x) \geq f(\lambda z + (1 - \lambda)x) \geq f(x) + \langle x^*, \lambda z + (1 - \lambda)x - x \rangle.$$

Esto nos dice que

$$f(z) \geq f(x) + \langle x^*, z - x \rangle.$$

Como esto es válido para cada $z \in X$, concluimos que $x^* \in \partial f(x)$. \square

Proposición 4.19. *Sea $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa. Si f es continua en x , entonces $\partial f(x)$ es acotado y no vacío.*

Demostración. De acuerdo con la Proposición 4.15, $\text{int}(\text{epi}(f)) \neq \emptyset$. Utilizamos la parte i) del Teorema de Hahn-Banach (Teorema 3.5) con $A = \text{int}(\text{epi}(f))$ y $B = \{(x, f(x))\}$. Luego, existen $L \in X^*$ y $s \in \mathbf{R}$, no ambos nulos, tales que

$$Ly + s\lambda \leq Lx + sf(x)$$

para todo $(y, \lambda) \in \text{int}(\text{epi}(f))$. Tomando $y = x$, concluimos que $s \leq 0$. Si $s = 0$ tendremos que $L(y - x) \leq 0$ para todo y en una vecindad de x . Esto nos dice que $L = 0$, lo cual es una contradicción. Concluimos que $s < 0$. Tenemos entonces que

$$\lambda \geq f(x) + \langle \ell, y - x \rangle,$$

con $\ell = -L/s$, para cada $\lambda > f(y)$. Pasando al límite, vemos que

$$f(y) \geq f(x) + \langle \ell, y - x \rangle.$$

La Proposición 4.18 nos dice que $\ell \in \partial f(x) \neq \emptyset$. Por otra parte, como f es continua en x , la Proposición 4.15 garantiza que ella es Lipschitz-continua en una vecindad de x . Si $x^* \in \partial f(x)$ tenemos que

$$f(x) + \langle x^*, y - x \rangle \leq f(y) \leq f(x) + M\|y - x\|,$$

y, por lo tanto, $\langle x^*, y - x \rangle \leq M\|y - x\|$ para todo y en una vecindad de x . Concluimos que $\|x^*\| \leq M$ y así, $\partial f(x) \subset \bar{B}(0, M)$. \square

Recordemos que la Proposición 4.5 establece, en particular, que si f es diferenciable en x , entonces f es subdiferenciable en dicho punto. La Proposición 4.19 va más allá, pues señala que basta que f sea continua en x para ser subdiferenciable allí.

Terminamos esta subsección con una importante regla de cálculo que establece condiciones suficientes para que el subdiferencial de la suma de dos funciones coincida con la suma de los subdiferenciales. Notemos primero que esto no siempre ocurre.

Ejemplo 4.20. Sean $f, g : \mathbf{R} \rightarrow \mathbf{R} \cup \{+\infty\}$ las funciones definidas por

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ +\infty & \text{si } x > 0 \end{cases} \quad \text{y} \quad g(x) = \begin{cases} +\infty & \text{si } x < 0 \\ \sqrt{x} & \text{si } x \geq 0. \end{cases}$$

Tenemos que

$$\partial f(x) = \begin{cases} \{0\} & \text{si } x < 0 \\ [0, +\infty) & \text{si } x = 0 \\ \emptyset & \text{si } x > 0 \end{cases} \quad \text{y} \quad \partial g(x) = \begin{cases} \emptyset & \text{si } x \leq 0 \\ \left\{ -\frac{1}{2\sqrt{x}} \right\} & \text{si } x > 0. \end{cases}$$

Por lo tanto,

$$\partial f(x) + \partial g(x) = \emptyset$$

para todo $x \in \mathbf{R}$. Por otra parte,

$$(f + g)(x) = \begin{cases} 0 & \text{si } x = 0 \\ +\infty & \text{si } x \neq 0, \end{cases}$$

con lo cual

$$\partial(f + g)(x) = \begin{cases} \mathbf{R} & \text{si } x = 0 \\ \emptyset & \text{si } x \neq 0. \end{cases}$$

El problema con las funciones del ejemplo anterior es, en términos simples, que sus dominios no se intersecan suficiente. Tenemos lo siguiente:

Teorema 4.21 (Teorema de Moreau-Rockafellar). *Sean $f, g : X \rightarrow \mathbf{R} \cup \{+\infty\}$ dos funciones convexas, propias e inferiormente semicontinuas. Para cada $x \in X$, se tiene que*

$$\partial f(x) + \partial g(x) \subset \partial(f + g)(x).$$

La igualdad se cumple para todo $x \in X$ si además f es continua en algún punto $x_0 \in \text{dom}(g)$.

Demostración. Si $x^* \in \partial f(x)$ y $z^* \in \partial g(x)$, entonces

$$\begin{aligned} f(y) &\geq f(x) + \langle x^*, y - x \rangle \\ g(y) &\geq g(x) + \langle z^*, y - x \rangle \end{aligned}$$

para cada $y \in X$. Sumando ambas desigualdades, concluimos que

$$f(y) + g(y) \geq f(x) + g(x) + \langle x^* + z^*, y - x \rangle$$

para todo $y \in X$ y, por lo tanto, $x^* + z^* \in \partial(f + g)(x)$.

Supongamos ahora que $u^* \in \partial(f + g)(x)$. Es decir,

$$f(y) + g(y) \geq f(x) + g(x) + \langle u^*, y - x \rangle \quad (4.4)$$

para todo $y \in X$. Encontraremos $x^* \in \partial f(x)$ y $z^* \in \partial g(x)$ tales que $x^* + z^* = u^*$. Definamos los conjuntos convexos

$$\begin{aligned} B &= \{(y, \lambda) \in X \times \mathbf{R} : g(y) - g(x) \leq -\lambda\} \\ C &= \{(y, \lambda) \in X \times \mathbf{R} : f(y) - \langle u^*, y - x \rangle - f(x) \leq \lambda\}. \end{aligned}$$

Definimos $h : X \rightarrow \mathbf{R} \cup \{+\infty\}$ como $h(y) = f(y) - \langle u^*, y - x \rangle - f(x)$. Luego, h es convexa, es continua en algún punto x_0 , y $C = \text{epi}(h)$. La Proposición 4.15 nos dice que el conjunto $A = \text{int}(C)$ es un abierto no vacío (también es convexo). Además $A \cap B = \emptyset$ en virtud de la desigualdad (4.4). El Teorema de Hahn-Banach (Teorema 3.5) nos garantiza la existencia de $(L, s) \in X^* \times \mathbf{R} \setminus \{(0, 0)\}$ tales que

$$\langle L, y \rangle + s\lambda \leq \langle L, z \rangle + s\mu$$

para cada $(y, \lambda) \in A$ y cada $(z, \mu) \in B$. En particular, tomando $(y, \lambda) = (x, 1) \in A$ y $(z, \mu) = (x, 0) \in B$, deducimos que $s \leq 0$. Por otra parte, si $s = 0$, tomando $z = x$ vemos que $\langle L, x - y \rangle \geq 0$ para todo y en una vecindad de x . Esto dice que $L = 0$, lo cual contradice el Teorema de Hahn-Banach pues $(L, s) \neq (0, 0)$. Por lo tanto, $s < 0$, y podemos escribir

$$\langle z^*, y \rangle + \lambda \leq \langle z^*, z \rangle + \mu$$

con $z^* = -L/s$. Tomando de nuevo $(z, \mu) = (x, 0) \in B$, y recordando que $C = \text{epi}(h)$, deducimos que

$$\langle z^*, y - x \rangle + h(y) \leq 0.$$

Usando la desigualdad (4.4) vemos que

$$g(y) \geq g(x) + \langle z^*, y - x \rangle$$

para todo $y \in \text{dom}(g)$, y, por lo tanto, para todo $y \in X$. Esto nos dice que $z^* \in \partial g(x)$. De manera similar mostramos que $x^* = u^* - z^* \in \partial f(x)$, de donde obtenemos el resultado. \square

Corolario 4.22. *Sea $C \subset X$ un conjunto convexo cerrado y no vacío. Supongamos que $f : X \rightarrow \mathbf{R}$ es una función propia, convexa y diferenciable. Entonces \bar{x} minimiza f sobre C si, y solo si,*

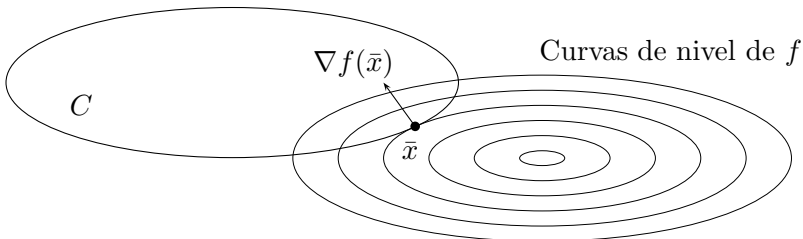
$$-Df(\bar{x}) \in N_C(\bar{x}).$$

Demostración. De acuerdo con la Regla de Fermat Generalizada (Teorema 4.7) y el Teorema de Moreau-Rockafellar (Teorema 4.21) tenemos que \bar{x} minimiza f sobre C si, y solo si,

$$0 \in \partial(f + \delta_C)(\bar{x}) = \partial f(\bar{x}) + \partial \delta_C(\bar{x}).$$

Como $\partial f(\bar{x}) = \{Df(\bar{x})\}$ y $\partial \delta_C(\bar{x}) = N_C(\bar{x})$, lo anterior es equivalente a $-Df(\bar{x}) \in N_C(\bar{x})$. \square

Para fijar ideas (pensar en $X = \mathbf{R}^N$), el corolario anterior nos dice que el gradiente de f en \bar{x} debe apuntar *hacia adentro* de C ; o, en otras palabras, que f solamente puede decrecer *hacia afuera* de C .



Capítulo 5

Ejemplos

5.1 Operador de Proximidad

Sea H un espacio de Hilbert.¹ Dados $\lambda > 0$ y $x \in H$, la *Regularizada de Moreau-Yosida* de f con parámetro (λ, x) es la función $f_{(\lambda, x)} : H \rightarrow \mathbf{R} \cup \{+\infty\}$, definida mediante

$$f_{(\lambda, x)}(z) = f(z) + \frac{1}{2\lambda} \|z - x\|^2.$$

Una propiedad importante es la siguiente:

Proposición 5.1. *Para cada (λ, x) , la función $f_{(\lambda, x)}$ tiene exactamente un minimizador \bar{x} , el cual está caracterizado por la relación*

$$-\frac{\bar{x} - x}{\lambda} \in \partial f(\bar{x}).$$

Demostración. Dado que f es propia, convexa e inferiormente semicontinua, tenemos que $f_{(\lambda, x)}$ es propia, estrictamente convexa, inferiormente semicontinua y coerciva. Obtenemos la existencia y unicidad gracias al Corolario 3.14. Finalmente, la Regla de Fermat Generalizada (Teorema 4.7) y el Teorema de Moreau-Rockafellar (Teorema 4.21) nos dicen que el único minimizador \bar{x} de la función $f_{(\lambda, x)}$ está caracterizado por

$$0 \in \partial f_{(\lambda, x)}(\bar{x}) = \partial f(\bar{x}) + \frac{\bar{x} - x}{\lambda},$$

lo que nos permite concluir. □

¹Las principales propiedades se presentan en el Apéndice A.

Utilizaremos este hecho en el Capítulo 8 cuando estudiemos el *algoritmo del punto proximal*.

Despejando \bar{x} , podemos escribir

$$\bar{x} = (I + \lambda \partial f)^{-1}x,$$

donde $I : H \rightarrow H$ denota la función identidad. En vista de la Proposición 5.1, la expresión $J_\lambda = (I + \lambda \partial f)^{-1}$ define una función $J_\lambda : H \rightarrow H$, que se llama *operador de proximidad* de f , con parámetro λ .²

Proposición 5.2. J_λ es Lipschitz-continua, con constante $M = 1$.

Demostración. Supongamos que $\bar{x} = J_\lambda(x)$ y $\bar{y} = J_\lambda(y)$, de manera que

$$-\frac{\bar{x} - x}{\lambda} \in \partial f(\bar{x}) \quad \text{y} \quad -\frac{\bar{y} - y}{\lambda} \in \partial f(\bar{y}).$$

De acuerdo con la Proposición 4.14, tenemos que

$$\langle (\bar{x} - x) - (\bar{y} - y), \bar{x} - \bar{y} \rangle \leq 0,$$

de donde $\|\bar{x} - \bar{y}\|^2 \leq \langle y - x, \bar{x} - \bar{y} \rangle \leq \|x - y\| \|\bar{x} - \bar{y}\|$. Concluimos que $\|\bar{x} - \bar{y}\| \leq \|x - y\|$. \square

Ejemplo 5.3. Si C es un subconjunto convexo, cerrado y no vacío de H , entonces la *función indicatriz* de C (ver Sección 2.1) es propia, convexa e inferiormente semicontinua. Dados $\lambda > 0$ y $x \in H$, $J_\lambda(x)$ es la única solución de

$$\min \left\{ \delta_C(z) + \frac{1}{2\lambda} \|z - x\|^2 : z \in H \right\} = \min \{ \|z - x\| : z \in C \}.$$

Independientemente de λ , $J_\lambda(x)$ es el punto de C que está más cerca de x . En otras palabras, es la *proyección ortogonal* de x sobre C . De acuerdo con la Proposición 5.2, la función $\text{Proy}_C : H \rightarrow H$, definida por $\text{Proy}_C(x) = J_\lambda(x)$, es Lipschitz-continua con constante $M = 1$.

²También se conoce como la *resolvente* del operador ∂f con parámetro λ .

5.2 Cálculo de Variaciones

Sea $L : [0, T] \times \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ una función continuamente diferenciable.

$$C = \left\{ x \in \mathcal{C}^1([0, T]; \mathbf{R}) : x(0) = x_0, x(T) = x_T \right\}.$$

Para $x \in C$ definamos

$$f(x) = \int_0^T L(t, x(t), \dot{x}(t)) dt$$

y consideremos el problema

$$(CV) \quad \min\{f(x) : x \in C\}.$$

La función f es diferenciable y

$$Df(x)h = \int_0^T \left[\partial_2 L(t, x(t), \dot{x}(t)) h(t) + \partial_3 L(t, x(t), \dot{x}(t)) \dot{h}(t) \right] dt.$$

Antes de presentar las condiciones necesarias de optimalidad, mostraremos el siguiente resultado auxiliar:

Lema 5.4. Sean $\alpha, \beta \in \mathcal{C}^0([0, T]; \mathbf{R})$ tales que

$$\int_0^T \left[\alpha(t)h(t)dt + \beta(t)\dot{h}(t) \right] dt = 0$$

para toda $h \in \mathcal{C}^1([0, T]; \mathbf{R})$ tal que $h(0) = h(T) = 0$. Entonces β es continuamente diferenciable y $\dot{\beta} = \alpha$.

Demostración. Analicemos primero el caso en que $\alpha \equiv 0$. Debemos probar que β es constante. Para ello definamos la función $H : [0, T] \rightarrow \mathbf{R}$ como

$$H(t) = \int_0^t [\beta(s) - B] ds, \quad \text{con} \quad B = \frac{1}{T} \int_0^T \beta(t) dt.$$

Claramente, H es continuamente diferenciable y $H(0) = H(T) = 0$. Luego

$$\begin{aligned} 0 &= \int_0^T \beta(t) \dot{H}(t) dt \\ &= \int_0^T \beta(t) [\beta(t) - B] dt \\ &= \int_0^T [\beta(t) - B]^2 dt + B \int_0^T [\beta(t) - B] dt \\ &= \int_0^T [\beta(t) - B]^2 dt. \end{aligned}$$

Como β es continua, tenemos que $\beta \equiv B$.

Pasemos ahora al caso general. Para ello, definamos

$$A(t) = \int_0^t \alpha(s) ds.$$

Usando integración por partes vemos que, para cada $h \in \mathcal{C}^1([0, T]; \mathbf{R})$ tal que $h(0) = h(T) = 0$, se cumple que

$$0 = \int_0^T [\alpha(t)h(t)dt + \beta(t)\dot{h}(t)] dt = \int_0^T [\beta(t) - A(t)] \dot{h}(t) dt.$$

De acuerdo con lo que demostramos antes, la función $\beta - A$ es constante. En otras palabras, β es una primitiva de α . Esto nos dice que β es continuamente diferenciable y $\dot{\beta} = \alpha$. \square

Estamos listos para presentar las condiciones necesarias de optimalidad:

Teorema 5.5 (Ecuación de Euler-Lagrange). *Sea x^* una solución de (CV). Entonces la función $\beta : [0, T] \rightarrow \mathbf{R}$, definida por*

$$\beta(t) = \partial_3 L(t, x^*(t), \dot{x}^*(t)),$$

es continuamente diferenciable y

$$\frac{d}{dt} \left[\partial_3 L(t, x^*(t), \dot{x}^*(t)) \right] = \partial_2 L(t, x^*(t), \dot{x}^*(t))$$

para todo $t \in (0, T)$.

Demostración. En el subespacio

$$C_0 = \left\{ x \in C^1([0, T]; \mathbf{R}) : x(0) = x(T) = 0 \right\},$$

de $C^1([0, T]; \mathbf{R})$, definamos la función $g : C_0 \rightarrow \mathbf{R}$ mediante

$$g(h) = f(x^* + h).$$

Claramente, tenemos que

$$\text{mín}\{f(x) : x \in C\} = \text{mín}\{g(h) : h \in C_0\}.$$

Sabemos que $0 \in \text{argmin}_{C_0}(g)$ pues $x^* \in \text{argmin}_C(f)$. Por otra parte, $Dg(0) = Df(x^*)$. El Teorema 4.2 nos dice que $Df(x^*)h = Dg(0)h = 0$ para toda $h \in C_0$. Es decir,

$$\int_0^T \left[\partial_2 L(t, x(t), \dot{x}(t))h(t) + \partial_3 L(t, x(t), \dot{x}(t))\dot{h}(t) \right] dt = 0$$

para toda $h \in C_0$. El resultado se obtiene utilizando el Lema 5.4 con $\alpha(t) = \partial_2 L(t, x^*(t), \dot{x}^*(t))$ y $\beta(t) = \partial_3 L(t, x^*(t), \dot{x}^*(t))$. \square

5.3 Algunas Ecuaciones Elípticas

Sea Ω un subconjunto abierto y acotado de \mathbf{R}^N y consideremos el espacio de Hilbert

$$X = H^1(\Omega; \mathbf{R}) = \{u \in L^2(\Omega; \mathbf{R}) : \nabla u \in L^2(\Omega; \mathbf{R}^N)\}^3$$

con el producto interno

$$\langle u, v \rangle = \int_{\Omega} \left[\nabla u(\xi) \cdot \nabla v(\xi) + u(\xi)v(\xi) \right] d\xi$$

y la norma asociada

$$\|u\| = \left[\int_{\Omega} \left[|\nabla u(\xi)|^2 + |u(\xi)|^2 \right] d\xi \right]^{1/2}.$$

³El gradiente se toma en el sentido de distribuciones.

Dada $h \in L^2(\Omega)$, definamos $f : X \rightarrow \mathbf{R}$ mediante

$$f(u) = \frac{1}{2}\|u\|^2 - \int_{\Omega} h(\xi)u(\xi) d\xi.$$

La función f es estrictamente convexa, coerciva y diferenciable con

$$Df(u)v = \langle u, v \rangle - \int_{\Omega} h(\xi)v(\xi) d\xi.$$

De acuerdo con el Corolario 3.14, f tiene un único minimizador \bar{u} en X y, gracias a la Regla de Fermat (Teorema 4.2), se cumple que

$$Df(\bar{u})v = \int_{\Omega} \left[\nabla \bar{u}(\xi) \cdot \nabla v(\xi) + \bar{u}(\xi)v(\xi) - h(\xi)v(\xi) \right] d\xi = 0 \quad (5.1)$$

para toda $v \in X$. Se puede demostrar (ver [5, Teorema 9.26]) que si Ω y h son suficientemente regulares, entonces $\bar{u} \in C^2(\bar{\Omega})$.⁴ Utilizando (5.1) y la Fórmula de Green⁵ vemos que

$$\int_{\Omega} \left[-\Delta \bar{u}(\xi) + u(\xi) \right] v(\xi) d\xi + \int_{\partial\Omega} \frac{\partial \bar{u}}{\partial n}(\xi)v(\xi) d\sigma(\xi) = \int_{\Omega} h(\xi)v(\xi) d\xi$$

para toda $v \in X$, en particular para aquellas que son continuas y se anulan en la frontera. Esto nos dice que

$$-\Delta \bar{u}(\xi) + u(\xi) = h(\xi)$$

para todo $\xi \in \Omega$ (o casi todo, según la regularidad de h). Luego,

$$\int_{\partial\Omega} \frac{\partial \bar{u}}{\partial n}(\xi)v(\xi) d\sigma(\xi) = 0$$

para toda $v \in X$ y, por lo tanto,

$$\frac{\partial \bar{u}}{\partial n}(\xi) = 0$$

en $\partial\Omega$. Hemos resuelto la ecuación con condición de Neumann:

$$\begin{cases} -\Delta u + u = h & \text{en } \Omega \\ \frac{\partial u}{\partial n} = 0 & \text{en } \partial\Omega. \end{cases}$$

Con argumentos similares se pueden resolver otras ecuaciones, por ejemplo, con condiciones de Dirichlet (ver [5, Capítulo 9]).

⁴Esto quiere decir que \bar{u} es dos veces continuamente diferenciable en Ω y la segunda derivada es uniformemente continua, por lo que \bar{u} , $\nabla \bar{u}$ y $\Delta \bar{u}$ se pueden extender continuamente a la frontera.

⁵ $\int_{\Omega} [(\Delta u)v + \nabla u \cdot \nabla v] = \int_{\partial\Omega} \frac{\partial u}{\partial n} v$, donde n denota la normal unitaria exterior a Ω .

Parte II

**Sistemas Dinámicos en
Minimización Convexa**

Capítulo 6

Sistemas Dinámicos

Lo que queda de este curso está dedicado al estudio de algunos sistemas dinámicos que permiten resolver problemas de minimización convexa en espacios de Hilbert.¹ En este capítulo entregaremos las principales definiciones, así como algunos ejemplos abstractos; mencionaremos los sistemas clásicos utilizados en minimización convexa; y presentaremos las principales herramientas que se utilizan para verificar la convergencia, fuerte o débil, de las trayectorias y sucesiones generadas a partir de estos sistemas.

6.1 Aspectos Generales

En términos muy generales, un sistema dinámico es una serie de reglas que rigen el movimiento de un objeto. En este curso nos interesaremos en dos tipos de sistemas:

Ecuaciones de Evolución: Son sistemas dinámicos a tiempo continuo, donde la transición está gobernada por una ecuación diferencial ordinaria o, más generalmente, una *inclusión diferencial*: Consideremos una función $F : [0, \infty) \times H \rightarrow 2^H$ (nótese que los *valores* de la función son subconjuntos de H). Una *solución* de la inclusión diferencial

$$\begin{cases} \dot{x}(t) \in F(t, x(t)), & t > 0 \\ x(0) = x_0 \end{cases}$$

¹Las principales propiedades se presentan en el Apéndice A.

es una función absolutamente continua $x : [0, \infty) \rightarrow H$ tal que $x(0) = x_0$ y $\dot{x}(t) \in F(t, x(t))$ para casi todo $t > 0$. Diremos que x_0 es la *condición inicial*, que $x(t)$ es el *estado del sistema en el instante t* , y que x es la *trayectoria que parte desde x_0* .

Si $F : [0, \infty) \times H \rightarrow H$ obtenemos una ecuación diferencial ordinaria:

$$\begin{cases} \dot{x}(t) &= F(t, x(t)), & t > 0 \\ x(0) &= x_0. \end{cases}$$

Algoritmos Iterativos: Son sistemas dinámicos a tiempo discreto, donde el estado del sistema en un tiempo dado se obtiene a partir del estado en el tiempo anterior al aplicar un procedimiento definido *a priori*. Más precisamente, partimos de una *condición inicial* $x_0 \in H$. En la *etapa* o *iteración* n , dado el estado x_n , determinamos el estado x_{n+1} mediante

$$x_{n+1} = F(n, x_n),$$

donde $F : \mathbf{N} \times H \rightarrow H$ es una función conocida de antemano.

6.2 Sistemas Dinámicos y Minimización

Comentaremos ahora algunos sistemas dinámicos básicos que pueden aplicarse para resolver problemas de minimización convexa. Para fijar ideas, comencemos por estudiar el caso diferenciable. Sea $f : H \rightarrow \mathbf{R}$ una función continuamente diferenciable, tal que ∇f es (globalmente) Lipschitz-continuo. Para cada $x_0 \in H$ la ecuación diferencial

$$(\mathcal{ED}) \quad \begin{cases} \dot{x}(t) &= -\nabla f(x(t)) \\ x(0) &= x_0 \end{cases}$$

tiene una única solución, en el sentido clásico; es decir, existe una función continuamente diferenciable $x : [0, \infty) \rightarrow H$ tal que $x(0) = x_0$ y $\dot{x}(t) = F(t, x(t))$ para todo $t > 0$.

En primer lugar, notemos que los puntos estacionarios de este sistema son exactamente los *puntos críticos* (aquellos donde el gradiente es cero) de f . Por otra parte, tenemos que

$$\frac{d}{dt}f(x(t)) = \langle \nabla f(x(t)), \dot{x}(t) \rangle = -\|\nabla f(x(t))\|^2 = -\|\dot{x}(t)\|^2.$$

Esto nos dice que los valores de f decrecen estrictamente a lo largo de la solución x , salvo en aquellos instantes en que el punto $x(t)$ es crítico. Finalmente, recordemos de la Observación 4.8, que si f es convexa, los únicos puntos críticos son los minimizadores globales. Intuitivamente, la trayectoria se moverá en direcciones que hagan decrecer el valor de f y no se detendrá a menos que se acerque a un minimizador global.

Si la función f no es diferenciable, la ecuación diferencial (\mathcal{ED}) deja de tener sentido. Sin embargo, si f es una función convexa, inferiormente semicontinua y propia, podemos considerar una versión más general de la ecuación diferencial (\mathcal{ED}) reemplazando ∇f por ∂f . Con esto se obtiene la inclusión diferencial

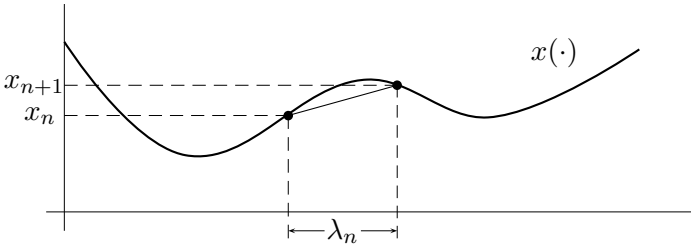
$$(\mathcal{ID}) \quad \begin{cases} -\dot{x}(t) \in \partial f(x(t)), & \text{para casi todo } t > 0 \\ x(0) = x_0. \end{cases}$$

En el Capítulo 7 presentaremos una noción de solución para (\mathcal{ID}) y enunciaremos el resultado de existencia y unicidad para cada condición inicial $x_0 \in \overline{\text{dom}(f)}$. Posteriormente, estudiaremos las principales propiedades de dichas soluciones y demostraremos que, cuando $t \rightarrow \infty$, cada una de ellas converge a algún minimizador de f . Esto nos da una manera, al menos teórica, de aproximar dichos puntos.

En los Capítulos 8 y 9 estudiaremos dos algoritmos que se definen, de manera muy intuitiva, a partir de (\mathcal{ED}) . Ellos nos darán métodos implementables, que nos permitirán aproximar los minimizadores de f desde un punto de vista práctico. La idea general es discretizar la ecuación diferencial, reemplazando la derivada por la “pendiente” de una secante. En otras palabras, aproximamos

$$\dot{x}(t) \sim \frac{x_{n+1} - x_n}{\lambda_n},$$

donde $\lambda_n > 0$ es el *paso de la discretización* (la longitud del n -ésimo subintervalo).



Para aproximar $\nabla f(x(t))$ hay dos opciones evidentes:

1. La primera es tomar $\nabla f(x_{n+1})$, que corresponde al valor del lado derecho del subintervalo. En este caso obtenemos una expresión implícita para x_{n+1} en función de x_n :

$$x_{n+1} + \lambda_n \nabla f(x_{n+1}) = x_n.$$

Para determinar x_{n+1} es necesario resolver una ecuación, lo que típicamente aumenta el costo computacional de cada iteración. Sin embargo, veremos que este algoritmo converge bajo hipótesis muy generales respecto de la función f . Este enfoque da origen al *algoritmo del punto proximal*, que se presenta en el Capítulo 8.

2. La segunda opción consiste en aproximar $\nabla f(x(t))$ por $\nabla f(x_n)$, que corresponde al valor del lado izquierdo del subintervalo. La ventaja de este enfoque es que se obtiene una fórmula explícita para calcular x_{n+1} en función de x_n :

$$x_{n+1} = x_n - \lambda_n \nabla f(x_n).$$

Sin embargo, solamente podremos demostrar la convergencia de este algoritmo bajo algunas hipótesis de regularidad sobre f y siempre que los pasos de la discretización no sean demasiado grandes. Estudiaremos el *algoritmo del gradiente*, basado en este enfoque, en el Capítulo 9.

6.3 Análisis Asintótico

En los capítulos que siguen estudiaremos el comportamiento asintótico de los sistemas dinámicos descritos arriba. Más precisamente, los Teoremas 7.3, 8.3 y 9.6 contienen los resultados de convergencia débil para

los tres sistemas, mientras que los Teoremas 7.4, 8.4 y 9.7 establecen la convergencia fuerte (cuando $t \rightarrow +\infty$ o $n \rightarrow +\infty$, según corresponda).

Veremos ahora algunos conceptos y herramientas que utilizaremos para comprobar la convergencia, débil o fuerte, a algún minimizador de la función objetivo.

6.3.1 Convergencia Débil

Presentaremos aquí un resultado muy útil para demostrar la convergencia débil, sin conocer el límite *a priori*.

Lema 6.1. *Sea F un subconjunto no vacío de un espacio de Hilbert H y sea $y : \mathbf{R}_+ \rightarrow H$. Supongamos que*

i) Para cada $u \in F$ existe $\lim_{t \rightarrow \infty} \|y(t) - u\|$; y

ii) Todo punto de acumulación débil de $y(t)$, con $t \rightarrow \infty$, está en F .

Entonces $y(t)$ converge débilmente cuando $t \rightarrow \infty$ a algún $\bar{u} \in F$.

Demostración. En vista de i), a partir de cierto T podemos suponer que $y(t)$ es acotada. En vista del Corolario 3.10, basta demostrar que no puede tener más de un punto de acumulación débil. Supongamos entonces que (t_n) y (s_n) son sucesiones tales que $\lim_{n \rightarrow \infty} t_n = \lim_{n \rightarrow \infty} s_n = \infty$, $y(t_n) \rightharpoonup x$ y $y(s_n) \rightharpoonup z$ cuando $n \rightarrow \infty$. Por ii), $x, z \in F$. Luego, por i), existen $\ell(x) = \lim_{t \rightarrow \infty} \|y(t) - x\|$ y $\ell(z) = \lim_{t \rightarrow \infty} \|y(t) - z\|$. Pero, dado que

$$\|y(t) - x\|^2 = \|y(t) - z\|^2 + \|z - x\|^2 + 2\langle y(t) - z, z - x \rangle,$$

pasando al límite en ambas subsucesiones obtenemos, respectivamente,

$$\ell(x) = \ell(z) - \|z - x\|^2 \quad \text{y} \quad \ell(x) = \ell(z) + \|z - x\|^2.$$

Concluimos que $z = x$. □

Dada una sucesión (y_n) en H podemos definir una interpolación constante por tramos, mediante

$$y(t) = y_n \quad \text{para} \quad t \in [n, n + 1).$$

Esto nos permite obtener el siguiente resultado, cuya demostración se deja como ejercicio.

Corolario 6.2 (Lema de Opial). *Sea F un subconjunto no vacío de un espacio de Hilbert H y sea (y_n) una sucesión en H . Supongamos que*

i) Para cada $u \in F$ existe $\lim_{n \rightarrow \infty} \|y_n - u\|$; y

ii) Todo punto de acumulación débil de (y_n) pertenece a F .

Entonces y_n converge débilmente cuando $n \rightarrow \infty$ a algún $\bar{u} \in F$.

6.3.2 Convergencia Fuerte

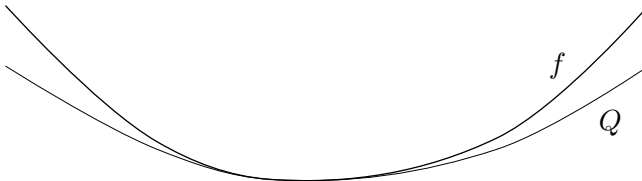
Para la convergencia fuerte se requieren hipótesis adicionales sobre la función objetivo. Comentaremos algunas de ellas:

Convexidad fuerte. Una función $f : H \rightarrow \mathbf{R} \cup \{+\infty\}$ es *fuertemente convexa* con parámetro $\alpha > 0$ si para cada $x \in \text{dom}(\partial f)$ y para cada $x^* \in \partial f(x)$ se cumple que

$$f(y) \geq f(x) + \langle x^*, y - x \rangle + \frac{\alpha}{2} \|y - x\|^2 \quad (6.1)$$

para todo $y \in H$. En otras palabras, esto quiere decir que f está por encima de la función cuadrática $Q : H \rightarrow \mathbf{R}$ definida por

$$Q(y) = f(x) + \langle x^*, y - x \rangle + \frac{\alpha}{2} \|y - x\|^2.$$



Toda función fuertemente convexa es a la vez estrictamente convexa y coerciva. Por lo tanto, el Corolario 3.14 nos dice que si f es propia, inferiormente semicontinua y fuertemente convexa, entonces debe tener exactamente un minimizador.

Paridad. Una función $f : H \rightarrow \mathbf{R} \cup \{+\infty\}$ es *par* si $f(-x) = f(x)$ para todo $x \in H$. Notemos que esto implica que $0 \in \operatorname{argmin}(f)$, pues

$$f(0) \leq \frac{1}{2}f(x) + \frac{1}{2}f(-x) = \frac{1}{2}f(x) + \frac{1}{2}f(x) = f(x)$$

para cada $x \in H$.

Conjunto de minimizadores con interior no vacío. Supongamos que $\operatorname{int}(\operatorname{argmin}(f)) \neq \emptyset$. Entonces, existen $\bar{u} \in \operatorname{argmin}(f)$ y $r > 0$ tales que $\bar{B}(\bar{u}, r) \subset \operatorname{argmin}(f)$. En particular, para cada $h \in H$ con $\|h\| = 1$ se tiene que $\bar{u} + rh \in \operatorname{argmin}(f)$. Dados $x \in \operatorname{dom}(\partial f)$ y $x^* \in \partial f(x)$, sabemos que

$$f(\bar{u} + rh) \geq f(x) + \langle x^*, \bar{u} + rh - x \rangle.$$

Como $\bar{u} + rh \in \operatorname{argmin}(f)$, lo anterior nos dice que

$$r\langle x^*, h \rangle \leq \langle x^*, x - \bar{u} \rangle.$$

En consecuencia,

$$r\|x^*\| = \sup_{\|h\|=1} \langle x^*, h \rangle \leq \langle x^*, x - \bar{u} \rangle. \quad (6.2)$$

Capítulo 7

Método del Máximo Descenso

El *método del máximo descenso* se basa en la inclusión diferencial

$$(\mathcal{ID}) \quad \begin{cases} -\dot{x}(t) \in \partial f(x(t)), & \text{para casi todo } t > 0 \\ x(0) = x_0. \end{cases}$$

Observemos que los puntos estacionarios para esta inclusión diferencial son precisamente los minimizadores de f .

Para cada condición inicial, existe una única *solución* de esta inclusión diferencial. Más precisamente, tenemos el siguiente:

Teorema 7.1. *Sea $f : H \rightarrow \mathbf{R} \cup \{+\infty\}$ una función convexa, inferiormente semicontinua y propia. Para cada $x_0 \in \overline{\text{dom}(f)}$ existe una única función absolutamente continua $x : [0, \infty) \rightarrow X$, que satisface la inclusión diferencial (\mathcal{ID}) .*

Existen dos demostraciones clásicas y bien conocidas de este resultado:

1. La primera se basa en un proceso de regularización: se sustituye la función f por una función continuamente diferenciable f_λ y se encuentra la solución x_λ de la ecuación diferencial

$$-\dot{x}_\lambda(t) = \nabla f_\lambda(x_\lambda(t)).$$

Luego, se prueba que cuando $\lambda \rightarrow 0$, la familia (x_λ) converge uniformemente en cualquier intervalo acotado a una trayectoria x que es solución de (\mathcal{ID}) . Este enfoque se puede encontrar en [4].

2. La segunda consiste en dividir un intervalo de la forma $[0, T]$ en n subintervalos y construir una función x_n (que pueden escogerse, por ejemplo, constante o lineal-afín en cada subintervalo) para encontrar la solución de (\mathcal{ID}) por aproximación. Este enfoque tiene su origen en [6] (ver también [8] y [11]).

Ambas estrategias tienen complejidades técnicas que preferimos omitir en este curso.

7.1 Propiedades del Sistema

Las principales propiedades de las soluciones de (\mathcal{ID}) se encuentran en la siguiente:

Proposición 7.2. *Sea $x : [0, \infty) \rightarrow X$ una solución de (\mathcal{ID}) .*

- i) La función $t \mapsto f(x(t))$ es decreciente.*
- ii) Para cada $u \in H$ se tiene que $\lim_{t \rightarrow \infty} f(x(t)) \leq f(u)$.¹ En particular, $\lim_{t \rightarrow \infty} f(x(t)) = \inf\{f(u) : u \in H\}$.*
- iii) Todo punto de acumulación débil de $x(t)$ cuando $t \rightarrow \infty$ debe pertenecer a $\operatorname{argmin}(f)$.*
- iv) Si $\bar{u} \in \operatorname{argmin}(f)$, la función $t \mapsto \|x(t) - \bar{u}\|$ es decreciente.*
- v) La trayectoria $x(\cdot)$ es acotada si, y solo si, $\operatorname{argmin}(f) \neq \emptyset$.*

Demostración. Dado que $-\dot{x}(t) \in \partial f(x(t))$, la definición de subgradiente nos dice que para todo $u \in H$ se cumple que

$$f(u) \geq f(x(t)) + \langle \dot{x}(t), x(t) - u \rangle. \quad (7.1)$$

Evaluando esta desigualdad en $u = x(s)$ con $s < t$ y dividiendo por $t - s$ deducimos que

$$\limsup_{s \rightarrow t^-} \frac{f(x(t)) - f(x(s))}{t - s} \leq -\|\dot{x}(t)\|^2 \leq 0,$$

¹Entendiéndose que el límite puede ser $-\infty$.

lo que nos da i). Por otra parte, como

$$\frac{1}{2} \frac{d}{dt} \|x(t) - u\|^2 = \langle \dot{x}(t), x(t) - u \rangle, \quad (7.2)$$

la desigualdad (7.1) nos dice que para cada $u \in H$ se tiene

$$f(u) \geq f(x(t)) + \frac{1}{2} \frac{d}{dt} \|x(t) - u\|^2.$$

Integrando entre 0 y T vemos, gracias a i), que

$$Tf(u) \geq Tf(x(T)) + \frac{1}{2} \|x(T) - u\|^2 - \frac{1}{2} \|x(0) - u\|^2.$$

Esto implica que

$$f(x(T)) \leq f(u) + \frac{1}{2T} \|x(0) - u\|^2,$$

de donde inmediatamente obtenemos ii). Para la parte iii), supongamos que $x(t_n)$ converge débilmente a \bar{x} . Como f es inferiormente semicontinua para la topología débil, tenemos que

$$f(\bar{x}) \leq \liminf_{n \rightarrow \infty} f(x(t_n)) = \lim_{t \rightarrow \infty} f(x(t)) \leq \inf\{f(u) : u \in H\},$$

con lo cual $\bar{x} \in \operatorname{argmin}(f)$. La parte iv) se obtiene utilizando (7.1) y (7.2) con $u = \bar{u} \in \operatorname{argmin}(f)$ pues

$$\frac{1}{2} \frac{d}{dt} \|x(t) - \bar{u}\|^2 = \langle \dot{x}(t), x(t) - \bar{u} \rangle \leq f(\bar{u}) - f(x(t)) \leq 0.$$

Finalmente, la parte v) se deduce de las partes iii) y iv), junto con el Teorema 3.9. \square

7.2 Convergencia

El siguiente resultado establece la convergencia débil de las trayectorias de (\mathcal{ID}) :

Teorema 7.3. *Sea $x : [0, \infty) \rightarrow X$ una solución de (\mathcal{ID}) y supongamos que $\operatorname{argmin}(f) \neq \emptyset$. Entonces $x(t)$ converge débilmente, cuando $t \rightarrow \infty$, a algún elemento de $\operatorname{argmin}(f)$.*

Demostración. Es directo a partir del Lema 6.1, en vista de los puntos iii) y iv) de la Proposición 7.2. \square

Para la convergencia fuerte tenemos el siguiente:

Teorema 7.4. *Sea $x : [0, \infty) \rightarrow X$ una solución de (ID) y supongamos que $\operatorname{argmin}(f) \neq \emptyset$. Entonces $x(t)$ converge fuertemente, cuando $t \rightarrow \infty$, a algún elemento de $\operatorname{argmin}(f)$, en cualquiera de los siguientes casos:*

- i) Si f es fuertemente convexa;
- ii) Si f es par; o
- iii) Si $\operatorname{int}(\operatorname{argmin}(f)) \neq \emptyset$.

Demostración. Supongamos primero que f es fuertemente convexa y tomemos $\bar{u} \in \operatorname{argmin}(f)$. Dado que $0 \in \partial f(\bar{u})$, la desigualdad (6.1), con $y = x(t)$, $x = \bar{u}$ y $x^* = 0$, nos dice que

$$\|x(t) - \bar{u}\|^2 \leq \frac{2}{\alpha} [f(x(t)) - f(\bar{u})].$$

Concluimos usando la parte ii) de la Proposición 7.2.

Supongamos ahora que f es par. Dados $t > 0$ y $s \in [0, t]$, definamos

$$\gamma(s) = \|u(s)\|^2 - \|u(t)\|^2 - \frac{1}{2}\|u(t) - u(s)\|^2.$$

Usando la parte i) de la Proposición 7.2, vemos que

$$\dot{\gamma}(s) = \langle \dot{x}(s), x(s) + x(t) \rangle \leq f(-x(t)) - f(x(s)) = f(x(t)) - f(x(s)) \leq 0.$$

Esto nos dice que γ es decreciente. Así, $\gamma(s) \geq \gamma(t) = 0$ y

$$\frac{1}{2}\|u(t) - u(s)\|^2 \leq \|u(s) - 0\|^2 - \|u(t) - 0\|^2.$$

Como $0 \in \operatorname{argmin}(f)$, la parte iv) de la Proposición 7.2 nos dice que $(x(t))_{t \geq 0}$ tiene la propiedad de Cauchy. Luego, es convergente porque el espacio H es completo.

Finalmente, supongamos que $\text{int}(\text{argmin}(f)) \neq \emptyset$. Entonces existen $\bar{u} \in \text{argmin}(f)$ y $r > 0$ tales que $\bar{B}(\bar{u}, r) \subset \text{argmin}(f)$. Utilizando la desigualdad (6.2) con $x = x(t)$ y $x^* = \dot{x}(t)$ vemos que

$$\begin{aligned}
 r\|x(t) - x(s)\| &= r \left\| \int_s^t \dot{x}(\tau) d\tau \right\| \\
 &\leq r \int_s^t \|\dot{x}(\tau)\| d\tau \\
 &\leq \int_s^t \langle \dot{x}(\tau), x(\tau) - \bar{u} \rangle d\tau \\
 &= \frac{1}{2} \int_s^t \frac{d}{d\tau} \|x(\tau) - \bar{u}\|^2 d\tau \\
 &= \frac{1}{2} \|x(t) - \bar{u}\|^2 - \frac{1}{2} \|x(s) - \bar{u}\|^2.
 \end{aligned}$$

La parte iv) de la Proposición 7.2 nos dice que $(x(t))_{t \geq 0}$ tiene la propiedad de Cauchy y, por lo tanto, converge. \square

Capítulo 8

Algoritmo del Punto Proximal

En lo que sigue, $f : H \rightarrow \mathbf{R} \cup \{+\infty\}$ es una función propia, convexa, e inferiormente semicontinua. Consideramos una sucesión (λ_n) de números positivos, siguiendo la notación de la Sección 6.2. Estudiaremos el *algoritmo del punto proximal*, propuesto en [9].

Algoritmo del Punto Proximal: Partimos de $x_0 \in H$. Para cada $n \geq 0$, dados el paso λ_n y el estado x_n , definimos el estado x_{n+1} como el único minimizador de la Regularizada de Moreau-Yosida $f_{(\lambda_n, x_n)}$, estudiada en la Sección 5.1. Es decir,

$$\{x_{n+1}\} = \operatorname{argmin} \left\{ f(z) + \frac{1}{2\lambda_n} \|z - x_n\|^2 \right\}. \quad (8.1)$$

El algoritmo está bien definido, gracias a la Proposición 5.1. Más aun,

$$0 \in \partial f_{(\lambda_n, x_n)}(x_{n+1}) = \partial f(x_{n+1}) + \frac{x_{n+1} - x_n}{\lambda_n}.$$

Es decir,

$$-\frac{x_{n+1} - x_n}{\lambda_n} \in \partial f(x_{n+1}). \quad (8.2)$$

Esto es precisamente una discretización de la (\mathcal{ID}) , como adelantamos en la Sección 6.2.

8.1 Propiedades del Sistema

Las principales propiedades de las sucesiones generadas por el algoritmo del punto proximal se resumen en la siguiente:

Proposición 8.1. *Sea (x_n) una sucesión generada por el algoritmo del punto proximal con $(\lambda_n) \notin \ell^1$. Tenemos lo siguiente:*

- i) *La sucesión $(f(x_n))$ es decreciente.*
- ii) *Para cada $u \in H$ se tiene que $\lim_{n \rightarrow \infty} f(x_n) \leq f(u)$.¹ En particular, $\lim_{n \rightarrow \infty} f(x_n) = \inf\{f(u) : u \in H\}$.*
- iii) *Todo punto de acumulación débil de la sucesión (x_n) debe pertenecer a $\operatorname{argmin}(f)$.*
- iv) *Si $x^* \in \operatorname{argmin}(f)$ entonces la sucesión $(\|x_n - x^*\|)$ es decreciente. En consecuencia, existe $\lim_{n \rightarrow \infty} \|x_n - x^*\|$.*
- v) *La sucesión (x_n) es acotada si, y solo si, $\operatorname{argmin}(f) \neq \emptyset$.*

Demostración. Una consecuencia directa de (8.1) es que para cada $n \geq 0$ se tiene que

$$f(x_{n+1}) + \frac{1}{2\lambda_n} \|x_{n+1} - x_n\|^2 \leq f(x_n),$$

lo que inmediatamente implica i). Si nos permitimos un pequeño abuso de terminología, podemos decir que $\lim_{n \rightarrow \infty} f(x_n)$ existe en $[-\infty, \infty)$. En el caso $\lim_{n \rightarrow \infty} f(x_n) = -\infty$, el punto ii) se cumple de manera trivial.

Por otra parte, la inclusión (8.2) y la definición del subdiferencial implican que para cada $u \in H$ se tiene que

$$f(u) \geq f(x_{n+1}) + \left\langle -\frac{x_{n+1} - x_n}{\lambda_n}, u - x_{n+1} \right\rangle.$$

Es decir,

$$\begin{aligned} 2\lambda_n [f(x_{n+1}) - f(u)] &\leq 2\langle x_{n+1} - x_n, u - x_{n+1} \rangle \\ &= \|x_n - u\|^2 - \|x_{n+1} - u\|^2 - \|x_{n+1} - x_n\|^2 \quad (8.3) \\ &\leq \|x_n - u\|^2 - \|x_{n+1} - u\|^2. \quad (8.4) \end{aligned}$$

¹Entendiéndose que el límite puede ser $-\infty$.

Sumando para $n = 0, \dots, N$ vemos que

$$2 \sum_{n=0}^N \lambda_n [f(x_{n+1}) - f(u)] \leq \|x_0 - u\|^2 - \|x_{N+1} - u\|^2 \leq \|x_0 - u\|^2$$

para cada $N \in \mathbf{N}$.

Si $\lim_{n \rightarrow \infty} f(x_n) > f(u)$ entonces existe $\delta > 0$ tal que $f(x_n) - f(u) \geq \delta$ para todo n suficientemente grande. Esto dice que $2\delta \sum_{n=0}^{\infty} \lambda_n < +\infty$, lo que contradice el que $(\lambda_n) \notin \ell^1$. Luego, $\lim_{n \rightarrow \infty} f(x_n) \leq f(u)$. Para la parte iii), supongamos que x_{k_n} converge débilmente a \bar{x} . Como f es inferiormente semicontinua para la topología débil, tenemos que

$$f(\bar{x}) \leq \liminf_{n \rightarrow \infty} f(x_{k_n}) = \lim_{n \rightarrow \infty} f(x_n) \leq \inf\{f(u) : u \in H\},$$

con lo cual $\bar{x} \in \operatorname{argmin}(f)$. El punto iv) se obtiene reemplazando $u = x^*$ en (8.4) pues $f(x_n) \geq f(x^*)$ para todo $n \in \mathbf{N}$. Finalmente, para probar v), notemos primero que si $\operatorname{argmin}(f) \neq \emptyset$, entonces la parte iv) implica que (x_n) tiene que ser acotada. Recíprocamente, si (x_n) es acotada, entonces tiene alguna subsucesión débilmente convergente. El límite débil debe pertenecer a $\operatorname{argmin}(f)$, en virtud de la parte iii). \square

Observación 8.2. Si $(\lambda_n) \in \ell^1$, las sucesiones generadas mediante el algoritmo del punto proximal siempre convergen fuertemente. Sin embargo, el límite puede no ser un minimizador de f (ver [10, Sección 3]).

8.2 Convergencia

El siguiente resultado establece la convergencia débil de las sucesiones generadas por el algoritmo del punto proximal:

Teorema 8.3. *Sea (x_n) una sucesión generada por el algoritmo del punto proximal con $(\lambda_n) \notin \ell^1$ y sea $\operatorname{argmin}(f) \neq \emptyset$. Entonces x_n converge débilmente, cuando $n \rightarrow \infty$, a algún elemento de $\operatorname{argmin}(f)$.*

Demostración. Basta combinar los puntos iii) y iv) de la Proposición 8.1 con el Lema de Opial (Corolario 6.2). \square

Para la convergencia fuerte, tenemos el siguiente:

Teorema 8.4. *Sea (x_n) una sucesión generada por el algoritmo del punto proximal con $(\lambda_n) \notin \ell^1$ y sea $\operatorname{argmin}(f) \neq \emptyset$. Entonces x_n converge fuertemente, cuando $n \rightarrow \infty$, a algún elemento de $\operatorname{argmin}(f)$, en cualquiera de los siguientes casos:*

- i) Si f es fuertemente convexa;*
- ii) Si f es par; o*
- iii) Si $\operatorname{int}(\operatorname{argmin}(f)) \neq \emptyset$.*

Demostración. Comencemos por suponer que f es fuertemente convexa y sea $\bar{u} \in \operatorname{argmin}(f)$. Como $0 \in \partial f(\bar{u})$, podemos utilizar la desigualdad (6.1), con $y = x_n$, $x = \bar{u}$ y $x^* = 0$, para deducir que

$$\|x_n - \bar{u}\|^2 \leq \frac{2}{\alpha} [f(x_n) - f(\bar{u})].$$

Concluimos usando la parte ii) de la Proposición 8.1.

Supongamos ahora que f es par. Tomemos $m > n$ y utilicemos la desigualdad (8.3) con $u = -x_m$ para obtener

$$\|x_{n+1} + x_m\|^2 \leq \|x_n + x_m\|^2.$$

Por lo tanto, para m fijo, la función $n \mapsto \|x_n + x_m\|^2$ es decreciente. En particular,

$$4\|x_m\|^2 = \|x_m + x_m\|^2 \leq \|x_n + x_m\|^2.$$

Por otra parte, la Identidad del Paralelogramo (Proposición A.6) nos da

$$\|x_n + x_m\|^2 + \|x_n - x_m\|^2 = 2\|x_m\|^2 + 2\|x_n\|^2.$$

Sumando obtenemos

$$\|x_n - x_m\|^2 \leq 2\|x_n\|^2 - 2\|x_m\|^2.$$

Como $0 \in \operatorname{argmin}(f)$, la parte iv) de la Proposición 8.1 nos dice que (x_n) es una sucesión de Cauchy y, por lo tanto, converge.

Finalmente, supongamos que $\operatorname{int}(\operatorname{argmin}(f)) \neq \emptyset$. Es decir, existen $\bar{u} \in \operatorname{argmin}(f)$ y $r > 0$ tales que $\bar{B}(\bar{u}, r) \subset \operatorname{argmin}(f)$.

Podemos usar la desigualdad (6.2) con $x = x_{n+1}$ y $x^* = \frac{x_n - x_{n+1}}{\lambda_n}$ para concluir que

$$\begin{aligned} 2r\|x_n - x_{n+1}\| &\leq 2\langle x_n - x_{n+1}, x_{n+1} - \bar{u} \rangle \\ &= \|x_n - \bar{u}\|^2 - \|x_{n+1} - \bar{u}\|^2 - \|x_{n+1} - x_n\|^2 \\ &\leq \|x_n - \bar{u}\|^2 - \|x_{n+1} - \bar{u}\|^2. \end{aligned}$$

De allí vemos que para cada $m > n$,

$$\begin{aligned} 2r\|x_n - x_m\| &\leq 2r \sum_{j=n}^{m-1} \|x_j - x_{j+1}\| \\ &\leq \|x_n - \bar{u}\|^2 - \|x_m - \bar{u}\|^2. \end{aligned}$$

Como antes, la parte iv) de la Proposición 8.1 nos dice que (x_n) es una sucesión de Cauchy y, por lo tanto, converge. \square

8.3 Consistencia de las Direcciones

Presentamos ahora otra propiedad importante de las sucesiones generadas utilizando este algoritmo. Mostraremos que las direcciones correspondientes a pasos sucesivos forman un ángulo agudo. En otras palabras, que el algoritmo “no se devuelve”. Esta consistencia entre las direcciones de avance del algoritmo explica, en parte, su gran estabilidad.

Proposición 8.5. *Supongamos que (x_n) es una sucesión generada mediante el algoritmo del punto proximal. Si $x_{n+1} \neq x_n$, entonces*

$$\langle x^{n+1} - x^n, x^n - x^{n-1} \rangle > 0.$$

Demostración. De acuerdo con (8.2), tenemos que

$$-\frac{x_{n+1} - x_n}{\lambda_n} \in \partial f(x_{n+1}) \quad \text{y} \quad -\frac{x_n - x_{n-1}}{\lambda_{n-1}} \in \partial f(x_n).$$

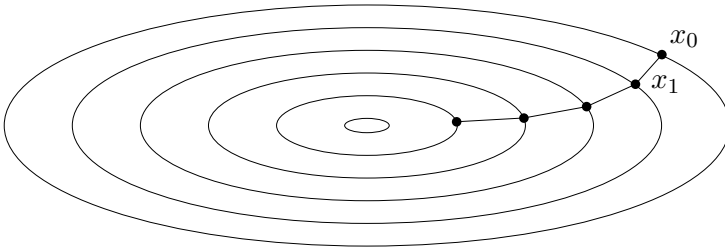
La Proposición 4.14 nos dice que

$$\left\langle \frac{x_{n+1} - x_n}{\lambda_n} - \frac{x_n - x_{n-1}}{\lambda_{n-1}}, x_{n+1} - x_n \right\rangle \leq 0.$$

Es decir,

$$\langle x_{n+1} - x_n, x_n - x_{n-1} \rangle \geq \frac{\lambda_{n-1}}{\lambda_n} \|x_{n+1} - x_n\|^2.$$

En particular, $\langle x^{n+1} - x^n, x^n - x^{n-1} \rangle > 0$ siempre que $x_{n+1} \neq x_n$. \square



Capítulo 9

Algoritmo del Gradiente

En esta sección consideraremos el *algoritmo del gradiente*, definido por

$$\begin{cases} z_{n+1} = z_n - \lambda_n \nabla f(z_n) \\ z_0 \in H. \end{cases}$$

Notemos que la expresión $z_{n+1} = z_n - \lambda_n \nabla f(z_n)$ es equivalente a

$$-\frac{z_{n+1} - z_n}{\lambda_n} = \nabla f(z_n),$$

con lo cual es una discretización explícita de la ecuación diferencial (\mathcal{ED}).

El análisis de convergencia para este algoritmo es más delicado que para el algoritmo del punto proximal. Comencemos por estudiar un caso sencillo en que el algoritmo del gradiente no converge.

Ejemplo 9.1. Consideremos la función $f : \mathbf{R} \rightarrow \mathbf{R}$, definida por $f(z) = z^2$. Tomemos $\lambda_n \equiv \lambda$ para todo n y veamos qué ocurre con el algoritmo partiendo desde el punto $z_0 = 1$. Es fácil ver que $z_n = (1 - 2\lambda)^n$ para cada n . Luego

- Si $\lambda = 1$, (z_n) se mantiene acotada pero no existe $\lim_{n \rightarrow \infty} z_n$.
- Si $\lambda > 1$, entonces $\lim_{n \rightarrow \infty} z_n = +\infty$.
- Si $\lambda < 1$, entonces $\lim_{n \rightarrow \infty} z_n = 0$, único minimizador de f .

Esto nos dice que para garantizar la convergencia de las sucesiones generadas por el algoritmo, tendremos que hacer alguna hipótesis adicional. Veremos que es suficiente suponer que la función es diferenciable con gradiente Lipschitz continuo, y que los pasos no son demasiado grandes.

9.1 Preliminares

Antes de dedicarnos a estudiar las propiedades del algoritmo, presentaremos dos resultados auxiliares que utilizaremos más adelante en el análisis:

Lema 9.2 (Lema de Descenso). *Si $f : H \rightarrow \mathbf{R}$ es continuamente diferenciable y ∇f es Lipschitz-continuo con constante L , entonces*

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

para cada $x, y \in H$.

Demostración. Escribamos $h = y - x$ y definamos $g : [0, 1] \rightarrow \mathbf{R}$ mediante $g(t) = f(x + th)$. La Regla de la Cadena nos da

$$\dot{g}(t) = \langle \nabla f(x + th), h \rangle,$$

de manera que

$$\int_0^1 \langle \nabla f(x + th), h \rangle dt = \int_0^1 \dot{g}(t) dt = g(1) - g(0) = f(y) - f(x).$$

Luego

$$\begin{aligned} f(y) - f(x) &\leq \int_0^1 \langle \nabla f(x), h \rangle dt + \int_0^1 \langle \nabla f(x + th) - \nabla f(x), h \rangle dt \\ &\leq \langle \nabla f(x), h \rangle + \int_0^1 \|\nabla f(x + th) - \nabla f(x)\| \|h\| dt \\ &\leq \langle \nabla f(x), h \rangle + L \|h\|^2 \int_0^1 t dt \\ &= \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2, \end{aligned}$$

como queríamos. □

Recordemos también el siguiente resultado sobre sucesiones reales:

Lema 9.3. Sean (a_n) y (ε_n) sucesiones de números no negativos tales que $(\varepsilon_n) \in \ell^1$ y $a_{n+1} - a_n \leq \varepsilon_n$ para cada $n \in \mathbf{N}$. Entonces existe $\lim_{n \rightarrow \infty} a_n$.

Demostración. Dado $n \geq 2$, definamos $w_n = a_n - \sum_{k=1}^{n-1} \varepsilon_k$. La sucesión (w_n) está acotada inferiormente y es decreciente. Luego, existe $\lim_{n \rightarrow \infty} w_n$. Así, también existe $\lim_{n \rightarrow \infty} a_n = \sum_{k=1}^{\infty} \varepsilon_k + \lim_{n \rightarrow \infty} w_n$. \square

9.2 Propiedades del Sistema

De acuerdo con el Ejemplo 9.1, es natural esperar que la convergencia de las sucesiones generadas por el algoritmo dependa de alguna hipótesis adicional que relacione los pasos de la discretización con el crecimiento de la función objetivo. Supondremos lo siguiente:

Hipótesis 9.4. La función ∇f es Lipschitz-continua con constante L y

$$\lambda_n \leq \lambda < \frac{2}{L} \quad \text{para todo } n \in \mathbf{N}.$$

Las sucesiones generadas por el algoritmo del gradiente tienen las siguientes propiedades:

Proposición 9.5. Supongamos que se cumple la Hipótesis 9.4 y sea (z_n) una sucesión generada por el algoritmo del gradiente con $(\lambda_n) \notin \ell^1$. Tenemos lo siguiente:

- i) La sucesión $(f(z_n))$ es decreciente.
- ii) Si $\inf(f) > -\infty$, la serie $\sum_{n=0}^{\infty} \|z_{n+1} - z_n\|^2$ es convergente.
- iii) Si $\inf(f) > -\infty$, entonces $\lim_{n \rightarrow \infty} f(z_n) = \inf(f)$ y todo punto de acumulación débil de la sucesión (z_n) debe pertenecer a $\operatorname{argmin}(f)$.
- iv) Si $\inf(f) > -\infty$ y (z_n) es acotada, entonces $\operatorname{argmin}(f) \neq \emptyset$.
- v) Si $z^* \in \operatorname{argmin}(f)$ entonces existe $\lim_{n \rightarrow \infty} \|z_n - z^*\|$. En particular, si $\operatorname{argmin}(f) \neq \emptyset$, entonces (z_n) es acotada.

Demostración. Del Lema 9.2 tenemos que

$$\begin{aligned} f(z_{n+1}) - f(z_n) &\leq \langle \nabla f(z_n), z_{n+1} - z_n \rangle + \frac{L}{2} \|z_{n+1} - z_n\|^2 \\ &= \left\langle \frac{z_n - z_{n+1}}{\lambda_n}, z_{n+1} - z_n \right\rangle + \frac{L}{2} \|z_{n+1} - z_n\|^2 \\ &= \left[\frac{L}{2} - \frac{1}{\lambda_n} \right] \|z_{n+1} - z_n\|^2. \end{aligned}$$

Teniendo en cuenta que $\lambda_n \leq \lambda < \frac{2}{L}$ para todo $n \in \mathbf{N}$ y escribiendo $\delta = \frac{2-\lambda L}{2\lambda} > 0$, concluimos que

$$f(z_{n+1}) - f(z_n) \leq -\delta \|z_{n+1} - z_n\|^2.$$

Los puntos i) y ii) son inmediatos, a partir de esta desigualdad. Para probar iii), de acuerdo con la definición del subdiferencial, deducimos que

$$f(u) \geq f(z_n) + \langle \nabla f(z_n), u - z_n \rangle = f(z_n) + \left\langle \frac{z_n - z_{n+1}}{\lambda_n}, u - z_n \right\rangle,$$

para cada $u \in H$. De allí,

$$\begin{aligned} 2\lambda_n [f(z_n) - f(u)] &\leq 2\langle z_{n+1} - z_n, u - z_n \rangle \\ &= \|z_n - u\|^2 - \|z_{n+1} - u\|^2 + \|z_{n+1} - z_n\|^2. \end{aligned} \quad (9.1)$$

Sumando para $n = 0, \dots, N$ vemos que

$$\begin{aligned} 2 \sum_{n=0}^N \lambda_n [f(z_n) - f(u)] &\leq \|z_0 - u\|^2 - \|z_{N+1} - u\|^2 + \sum_{n=0}^N \|z_{n+1} - z_n\|^2 \\ &\leq \|z_0 - u\|^2 + \sum_{n=0}^{\infty} \|z_{n+1} - z_n\|^2. \end{aligned}$$

Como $\inf(f) > -\infty$, la parte i) nos dice que existe $\lim_{n \rightarrow \infty} f(z_n)$. Dado que $(\lambda_n) \notin \ell^1$, deducimos que $\lim_{n \rightarrow \infty} f(z_n) = \inf(f)$, siguiendo un razonamiento análogo al de la demostración de la parte ii) de la Proposición 8.1. Después usamos la parte i) de la Proposición 3.12. El punto iv) se obtiene de la parte anterior extrayendo una subsucesión débilmente convergente. Finalmente, para verificar v), sustituimos $u = x^*$ en (9.1) y usamos el Lema 9.3 con $a_n = \|z_n - z^*\|$ y $\varepsilon_n = \|z_{n+1} - z_n\|^2$, que es sumable en virtud del punto ii). \square

9.3 Convergencia

Ahora estamos listos para establecer el resultado de convergencia débil para el algoritmo del gradiente:

Teorema 9.6. *Supongamos que $\operatorname{argmin}(f) \neq \emptyset$ y que se cumple la Hipótesis 9.4. Entonces cada sucesión $(z_n)_{n \in \mathbf{N}}$ generada por el algoritmo del gradiente con $(\lambda_n) \notin \ell^1$ converge débilmente, cuando $n \rightarrow \infty$, a algún elemento de $\operatorname{argmin}(f)$.*

Demostración. Basta combinar los puntos iii) y v) de la Proposición 9.5 con el Lema de Opial (Corolario 6.2). \square

Para la convergencia fuerte tenemos:

Teorema 9.7. *Supongamos que $\operatorname{argmin}(f) \neq \emptyset$ y que se cumple la Hipótesis 9.4. Entonces cada sucesión $(z_n)_{n \in \mathbf{N}}$ generada por el algoritmo del gradiente con $(\lambda_n) \notin \ell^1$ converge fuertemente, cuando $n \rightarrow \infty$, a algún elemento de $\operatorname{argmin}(f)$, en cualquiera de los siguientes casos:*

- i) *Si f es fuertemente convexa;*
- ii) *Si f es par; o*
- iii) *Si $\operatorname{int}(\operatorname{argmin}(f)) \neq \emptyset$.*

Demostración. En primer lugar, supongamos que f es fuertemente convexa y sea $\bar{u} \in \operatorname{argmin}(f)$. Como $0 \in \partial f(\bar{u})$, podemos utilizar la desigualdad (6.1), con $y = z_n$, $x = \bar{u}$ y $x^* = 0$, para deducir que

$$\|z_n - \bar{u}\|^2 \leq \frac{2}{\alpha} [f(z_n) - f(\bar{u})].$$

Concluimos usando la parte iii) de la Proposición 9.5.

Supongamos ahora que f es par. Tomemos $m > n$ y utilicemos la desigualdad (9.1) con $u = -z_m$ para obtener

$$\|z_{n+1} + z_m\|^2 \leq \|z_n + z_m\|^2 + \|z_{n+1} + z_n\|^2,$$

en vista de la parte i) de la Proposición 9.5. De allí,

$$4\|z_m\|^2 = \|z_m + z_m\|^2 \leq \|z_n + z_m\|^2 + \sum_{j=n}^{m-1} \|z_{j+1} + z_j\|^2.$$

Por otra parte, la Identidad del Paralelogramo (Proposición A.6) nos da

$$\|z_n + z_m\|^2 + \|z_n - z_m\|^2 = 2\|z_m\|^2 + 2\|z_n\|^2.$$

Sumando obtenemos

$$\|z_n - z_m\|^2 \leq 2\|z_n\|^2 - 2\|z_m\|^2 + \sum_{j=n}^{\infty} \|z_{j+1} + z_j\|^2,$$

por la parte ii) de la Proposición 9.5. Como $0 \in \operatorname{argmin}(f)$, la parte iv) de la Proposición 8.1 nos dice, como antes, que (z_n) es una sucesión de Cauchy.

Finalmente, supongamos que $\operatorname{int}(\operatorname{argmin}(f)) \neq \emptyset$. Es decir, existen $\bar{u} \in \operatorname{argmin}(f)$ y $r > 0$ tales que $\bar{B}(\bar{u}, r) \subset \operatorname{argmin}(f)$. Usamos la desigualdad (6.2) con $x = z_n$ y $x^* = \frac{z_n - z_{n+1}}{\lambda_n}$ para concluir que

$$\begin{aligned} 2r\|z_n - z_{n+1}\| &\leq 2\langle z_n - z_{n+1}, z_n - \bar{u} \rangle \\ &= \|z_n - \bar{u}\|^2 - \|z_{n+1} - \bar{u}\|^2 + \|z_{n+1} - z_n\|^2. \end{aligned}$$

De allí vemos que para cada $m > n$,

$$\begin{aligned} 2r\|z_n - z_m\| &\leq 2r \sum_{j=n}^{m-1} \|z_j - z_{j+1}\| \\ &\leq \|z_n - \bar{u}\|^2 - \|z_m - \bar{u}\|^2 + \sum_{j=n}^{\infty} \|z_j - z_{j+1}\|^2. \end{aligned}$$

Como antes, la parte v) de la Proposición 9.5 nos dice que (z_n) es una sucesión de Cauchy y, por lo tanto, converge. \square

Capítulo 10

Comentarios Finales

Comentaremos brevemente algunas relaciones entre los sistemas descritos en los capítulos precedentes. En primer lugar, veremos cómo se pueden aplicar al caso sencillo de una función cuadrática y destacaremos algunos aspectos positivos y negativos de cada método. Para terminar, mencionaremos cómo se pueden combinar los algoritmos del punto proximal y del gradiente para obtener un método de descomposición cuando la función objetivo se expresa como suma de dos funciones.

10.1 Caso de una Función Cuadrática

Consideremos una *función cuadrática* $f : H \rightarrow \mathbf{R}$ definida mediante

$$f(x) = \frac{1}{2}\langle x, Ax \rangle - \langle b, x \rangle + c,$$

donde $A : H \rightarrow H$ es un operador lineal y continuo, $b \in H$ y $c \in \mathbf{R}$. Suponemos además que A es simétrico (autoadjunto, por ser continuo) y semidefinido positivo. La función f es convexa y diferenciable con

$$\nabla f(x) = Ax - b.$$

Más aun, la función ∇f es Lipschitz-continua con constante

$$M = \|A\| = \sup_{\|x\|=1} \|Ax\|.$$

Como f es convexa, sus puntos críticos son, a la vez, sus minimizadores globales. Es decir,

$$\operatorname{argmin}(f) = \{x \in H : Ax = b\}.$$

10.1.1 Método del Máximo Descenso

La ecuación diferencial $(\mathcal{ED})^1$ es

$$\dot{x}(t) + Ax(t) = b, \quad (10.1)$$

cuyas soluciones (una para cada condición inicial x_0) existen y están definidas en todo \mathbf{R} . Si la ecuación $Ax = b$ no tiene solución, la parte v) de la Proposición 7.2 nos dice que las soluciones de la ecuación (10.1) no son acotadas. Si, por el contrario, $A\bar{x} = b$ para algún $\bar{x} \in H$, el Teorema 7.3 nos dice que cada solución $x(\cdot)$ de (10.1) converge débilmente a algún \hat{x} tal que $A\hat{x} = b$. Sin embargo, podemos hacer el cambio de variable $u(t) = x(t) - \bar{x}$ en (10.1) y obtenemos

$$\dot{u}(t) + Au(t) = 0.$$

Esta ecuación se puede escribir como

$$-\dot{u}(t) = \nabla g(u(t)),$$

con $g(x) = \frac{1}{2}\langle x, Ax \rangle$. Como la función g es par, el Teorema 7.4 nos dice que $u(t)$ converge fuertemente cuando $t \rightarrow \infty$ a algún \tilde{x} que satisface $0 = \nabla g(\tilde{x}) = A\tilde{x}$. Deducimos que $x(t)$ converge fuertemente a $\hat{x} = \bar{x} + \tilde{x}$, que, en efecto, es solución de $Ax = b$.

10.1.2 Algoritmo del Punto Proximal

El algoritmo del punto proximal queda descrito por

$$\begin{cases} x_0 \in H \\ x_{n+1} = (I + \lambda_n A)^{-1}(x_n + \lambda_n b), \quad n \geq 0. \end{cases}$$

Al igual que antes, la parte v) de la Proposición 8.1 nos dice que si $Ax = b$ no tiene solución, entonces la sucesión (x_n) no es acotada. En

¹No es necesario considerar la inclusión diferencial (\mathcal{ID}) pues f es diferenciable.

caso contrario, de acuerdo con el Teorema 8.3 converge débilmente a algún \hat{x} tal que $A\hat{x} = b$, siempre que $(\lambda_n) \notin \ell^1$. La convergencia fuerte es un tema más delicado (ver, por ejemplo, [1]).

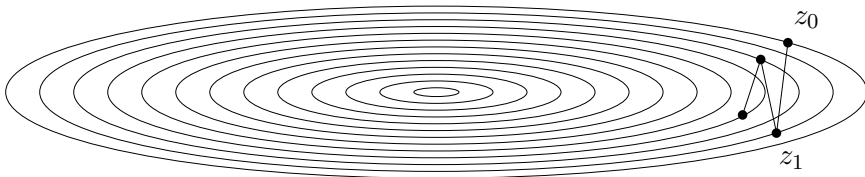
Si bien la convergencia del algoritmo ocurre bajo hipótesis muy generales, su principal desventaja es que para calcular x_{n+1} a partir de x_n es necesario invertir el operador $I + \lambda_n A$. Esto supone un costo computacional elevado y, posiblemente, la aparición de errores de cálculo, especialmente cuando la dimensión del espacio es grande o infinita.

10.1.3 Algoritmo del Gradiente

El algoritmo del gradiente queda descrito por

$$\begin{cases} z_0 \in H \\ z_{n+1} = z_n - \lambda_n(Az_n - b), \quad n \geq 0. \end{cases}$$

La convergencia débil está dada por el Teorema 8.3 si $(\lambda_n) \notin \ell^1$ y $\lambda_n \leq \lambda < \frac{2}{M}$ para todo $n \in \mathbf{N}$. Si M es muy grande, es necesario tomar λ_n muy pequeño. En algunos casos, esto provoca que el algoritmo avance muy lentamente. Otro factor que puede afectar la rapidez de convergencia de este algoritmo es el efecto *zigzag*, que ya hemos encontrado en el Ejemplo 9.1. Este fenómeno puede afectar especialmente al sistema cuando λ_n es muy cercano a $\frac{2}{M}$. La convergencia también es lenta cuando A es una matriz cuyos autovalores tienen escalas muy distintas (en \mathbf{R}^2 , cuando las curvas de nivel son elipses muy “aplastadas”).



Por otra parte, un punto a favor es que la fórmula para calcular z_{n+1} a partir de z_n es explícita y, por lo tanto, fácil de implementar.

10.2 Estrategias de Descomposición

Supongamos que queremos encontrar algún minimizador de una función $f : H \rightarrow \mathbf{R} \cup \{+\infty\}$, que se descompone como $f = g + h$, donde g es convexa, propia e inferiormente semicontinua, mientras que h es convexa, diferenciable y con gradiente Lipschitz-continuo. Es posible pensar en un algoritmo mixto, que combine una subiteración de tipo gradiente para la función que es regular, y otra de tipo proximal para la función que no es diferenciable. Esto permite aprovechar la regularidad para disminuir el costo computacional de cada iteración. Más precisamente, dado x_n , calculamos

$$w_n = x_n - \lambda_n \nabla h(x_n)$$

(subiteración de tipo gradiente) y luego determinamos x_{n+1} resolviendo

$$x_{n+1} + \lambda_n \partial g(x_{n+1}) \ni w_n$$

(subiteración de tipo proximal). En otras palabras, podemos escribir

$$x_{n+1} + \lambda_n \partial g(x_{n+1}) \ni x_n - \lambda_n \nabla h(x_n),$$

o, de manera equivalente,

$$\{x_{n+1}\} = \operatorname{argmin} \left\{ g(z) + \frac{1}{2\lambda_n} \|x_n - \lambda_n \nabla h(x_n) - z\|^2 \right\}.$$

Dejamos como ejercicio encontrar condiciones bajo las cuales las sucesiones generadas por este algoritmo mixto convergen, fuerte o débilmente, a algún minimizador de f .

Un caso especial es cuando $g = \delta_C$ es la función indicatriz de un conjunto convexo, cerrado y no vacío, $C \subset H$. Se obtiene el *algoritmo del gradiente proyectado*, cuyas iteraciones están definidas por

$$x_{n+1} = \operatorname{Proy}_C \left(x_n - \lambda_n \nabla h(x_n) \right),$$

en vista del Ejemplo 5.3.

Apéndice A

Espacios de Hilbert

A continuación presentamos algunos conceptos y resultados importantes relacionados con los espacios de Hilbert. No se pretende hacer una exposición exhaustiva, sino simplemente recordar aquellos puntos que son necesarios para seguir este curso.

A.1 Conceptos Básicos

Un *producto interno* en un espacio vectorial real H es una función $\langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbf{R}$ que es

Simétrica:

$$\langle x, y \rangle = \langle y, x \rangle \text{ para cada } x, y \in H;$$

Sesquilineal (lineal en la primera componente):

$$\langle \alpha x + y, z \rangle = \alpha \langle x, z \rangle + \langle y, z \rangle \text{ para cada } \alpha \in \mathbf{R} \text{ y } x, y, z \in H;$$

Definida positiva:

$$\langle x, x \rangle \geq 0 \text{ para todo } x \in H \text{ y } \langle x, x \rangle = 0 \text{ si, y solo si, } x = 0.$$

Observación A.1. En virtud de la simetría, el producto interno también es lineal en la segunda componente.

Las propiedades del producto interno nos permiten definir una función $\| \cdot \| : H \rightarrow \mathbf{R}$ mediante $\|x\| = \sqrt{\langle x, x \rangle}$. Claramente, $\|x\| \geq 0$ para todo

$x \in H$ y $\|x\| = 0$ si, y solo si, $x = 0$. Además, para cada $\alpha \in \mathbf{R}$ y $x \in H$ se tiene que $\|\alpha x\| = |\alpha|\|x\|$. Para comprobar que esta función es una norma en H , basta verificar que cumple la desigualdad triangular.

Proposición A.2. *Para cada $x, y \in H$ se cumple*

i) *La Desigualdad de Cauchy-Schwartz: $\langle x, y \rangle \leq \|x\| \|y\|$.*

ii) *La Desigualdad Triangular: $\|x + y\| \leq \|x\| + \|y\|$.*

Demostración. Claramente la Desigualdad de Cauchy-Schwartz es cierta si $y = 0$. Si $y \neq 0$ notemos que para cada $\alpha > 0$ se cumple que

$$0 \leq \|x - \alpha y\|^2 = \|x\|^2 - 2\alpha \langle x, y \rangle + \alpha^2 \|y\|^2.$$

Por lo tanto,

$$\langle x, y \rangle \leq \frac{1}{2\alpha} \|x\|^2 + \frac{\alpha}{2} \|y\|^2$$

para cada $\alpha > 0$. En particular, tomando $\alpha = \|x\|/\|y\|$, vemos que

$$\langle x, y \rangle \leq \frac{1}{2} \|x\| \|y\| + \frac{1}{2} \|x\| \|y\| = \|x\| \|y\|,$$

lo que demuestra i). Finalmente, utilizamos i) para deducir que

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2, \end{aligned}$$

de donde obtenemos ii). □

Decimos entonces que $\|\cdot\|$ es la *norma asociada* al producto interno $\langle \cdot, \cdot \rangle$. De aquí en adelante, cada espacio donde esté definido un producto interno, será para nosotros un espacio vectorial normado (con la norma asociada a dicho producto interno).

Un *espacio de Hilbert real* es un espacio vectorial sobre \mathbf{R} , donde se ha definido un producto interno, y que es completo para la topología definida por la norma asociada. Todo espacio de Hilbert es de Banach.

Ejemplo A.3. El espacio euclideo \mathbf{R}^N con el producto escalar.

Ejemplo A.4. El espacio de las sucesiones a cuadrado sumable $\ell^2(\mathbf{N}; \mathbf{R})$ con el producto interno $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{n \in \mathbf{N}} a_n b_n$.

Ejemplo A.5. El espacio de las funciones medibles a cuadrado integrable $L^2(\Omega; \mathbf{R})$ con el producto interno $\langle \phi, \psi \rangle = \int_{\Omega} \phi(\zeta) \psi(\zeta) d\zeta$.

Otra propiedad geométrica importante de los espacios de Hilbert viene dada por la siguiente:

Proposición A.6 (Identidad del Paralelogramo). *Para cada $x, y \in H$ se cumple que*

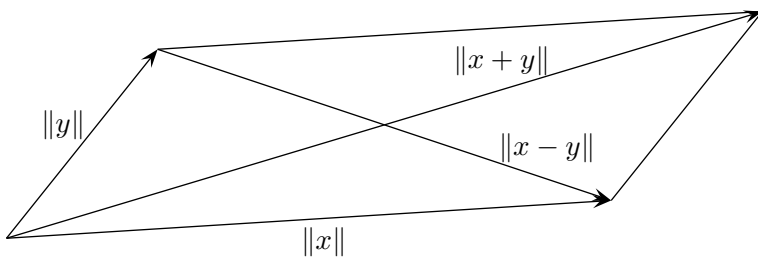
$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

Demostración. Tenemos que

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle \quad \text{y} \quad \|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle.$$

Basta sumar ambas igualdades para obtener el resultado. \square

La Proposición A.6 entrega una relación entre las longitudes de los lados y las diagonales de un paralelogramo.



También se puede demostrar que, si en un espacio de Banach se cumple la identidad del paralelogramo, entonces se puede definir un producto interno de manera que la norma asociada es precisamente la norma original. Esto se logra mediante la llamada *Identidad de Polarización*:

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

A.2 Proyección Ortogonal y Dualidad

Proposición A.7. *Sea K un subconjunto convexo, cerrado y no vacío de H y sea $x \in H$. Entonces existe un único $P_K(x) \in K$ tal que*

$$\|x - P_K(x)\| = \min_{y \in K} \|x - y\|.$$

Demostración. Sea $d = \min_{y \in K} \|x - y\|$ y consideremos una sucesión $(y_n) \in K$ tal que $\lim_{n \rightarrow \infty} \|y_n - x\| = d$. Tenemos que

$$\begin{aligned} \|y_n - y_m\|^2 &= \|(y_n - x) + (x - y_m)\|^2 \\ &= 2 \left(\|y_n - x\|^2 + \|y_m - x\|^2 \right) - \|(y_n + y_m) - 2x\|^2, \end{aligned}$$

en virtud de la Identidad del Paralelogramo (Proposición A.6). Como K es convexo, tenemos que $\frac{y_n + y_m}{2} \in K$, con lo cual

$$\|(y_n + y_m) - 2x\|^2 = 4 \left\| \frac{y_n + y_m}{2} - x \right\|^2 \geq 4d^2,$$

de acuerdo con la definición de d . Deducimos que

$$\|y_n - y_m\|^2 \leq 2 \left(\|y_n - x\|^2 + \|y_m - x\|^2 - 2d^2 \right).$$

Esto nos dice que la sucesión (y_n) es de Cauchy y por lo tanto converge a algún y^* . Dado que K es cerrado, $y^* \in K$. Finalmente, como la norma es continua, concluimos que $d = \lim_{n \rightarrow \infty} \|y_n - x\| = \|y^* - x\|$, por lo que basta definir $P_K(x) = y^*$. Para ver la unicidad, supongamos que $z^* \in K$ es tal que $\|x - z^*\| = \|x - P_K(x)\| = d$. La Proposición A.6 nos dice que

$$\|P_K(x) - z^*\|^2 = 4d^2 - \|P_K(x) + z^* - 2x\|^2.$$

Como K es convexo, tenemos que $\frac{P_K(x) + z^*}{2} \in K$. Luego,

$$\|P_K(x) + z^* - 2x\|^2 = 4 \left\| \frac{P_K(x) + z^*}{2} - x \right\|^2 \geq 4d^2,$$

y concluimos que $z^* = P_K(x)$. □

Con la notación de la Proposición A.7, decimos que $P_K(x)$ es la *proyección ortogonal de x sobre K* .

Proposición A.8. *Sea M un subespacio vectorial cerrado de H . Entonces,*

$$\langle x_0 - P_M(x_0), u \rangle = 0$$

para cada $x_0 \in H$ y cada $u \in M$.

Demostración. Sea $x_0 \in H$ y definamos $f : H \rightarrow \mathbf{R} \cup \{+\infty\}$ por

$$f(h) = \frac{1}{2} \|h - x_0\|^2 + \delta_M(h).$$

Como f es convexa y $P_M(x_0)$ minimiza f , la Regla de Fermat Generalizada (Teorema 4.7) nos dice que

$$0 \in \partial f(P_M(x_0)) = P_M(x_0) - x_0 + N_M(P_M(x_0)),$$

donde $N_M(z)$ denota el cono normal a M en z . Es decir,

$$\langle x_0 - P_M(x_0), h - P_M(x_0) \rangle \leq 0$$

para todo $h \in M$. Como M es subespacio vectorial y $P_M(x_0) \in M$, esto es equivalente a que

$$\langle x_0 - P_M(x_0), u \rangle = 0$$

para cada $u \in M$. □

Si $\langle x, y \rangle = 0$, decimos que x e y son *ortogonales* y escribimos $x \perp y$.

Dado $y \in H$, podemos definir $L_y : H \rightarrow \mathbf{R}$ mediante $L_y(h) = \langle y, h \rangle$. Claramente, la función L_y es lineal. También es continua, en virtud de la Desigualdad de Cauchy-Schwartz. Así, cada elemento de H define un elemento del dual topológico H^* . El recíproco también es cierto:

Teorema A.9 (Teorema de Representación de Riesz). *Sea H un espacio de Hilbert y sea $L \in H^*$ un funcional lineal continuo en H . Entonces existe $y_L \in H$ tal que*

$$L(h) = \langle y_L, h \rangle$$

para todo $h \in H$.

Demostración. Sea $M = \text{Ker}(L)$. Si $M = H$, entonces $L(h) = 0$ para todo $h \in H$ y basta tomar $y_L = 0$. Si $M \neq H$, escojamos cualquier $x_0 \notin M$ y definamos

$$x_1 = x_0 - P_M(x_0).$$

Observemos que $x_1 \neq 0$ y $x_1 \notin M$. Dado cualquier $h \in H$, definamos $u = L(x_1)h - L(h)x_1$, de modo que $u \in M$. Por la Proposición A.8, tenemos que $\langle x_1, u \rangle = 0$. Es decir,

$$0 = \langle x_1, u \rangle = \langle x_1, L(x_1)h - L(h)x_1 \rangle = L(x_1)\langle x_1, h \rangle - L(h)\|x_1\|^2.$$

Basta definir

$$y_L = \frac{L(x_1)}{\|x_1\|^2} x_1$$

para obtener el resultado. \square

Por medio del Teorema de Representación de Riesz (Teorema A.9) se puede identificar el dual topológico H^* con el propio H , mediante la función $\Phi : H \rightarrow H^*$, definida por $\Phi(y) = L_y$, la cual es un isomorfismo isométrico. Cuidando algunos detalles teóricos, esto además implica que todo espacio de Hilbert es, a su vez, un espacio de Banach reflexivo.

Índice alfabético

- algoritmo
 - del gradiente, 46, 63
 - del gradiente proyectado, 72
 - del punto proximal, 46, 57
 - iterativo, 44
 - mixto, 72
- argumento del mínimo, 1
- condición inicial, 44
- conjunto
 - convexo, 8
 - de subnivel, 8
 - factible, 1
- cono normal, 25
- convergencia
 - débil, 17
 - fuerte, 17
- desigualdad
 - de Cauchy-Schwartz, 74
 - triangular, 74
- dominio
 - de una función, 7
 - del subdiferencial, 23
- dual topológico, 15
- ecuación
 - de Euler-Lagrange, 38
 - de evolución, 43
- efecto zigzag, 71
- epígrafo, 8
- espacio
 - de Banach reflexivo, 17
 - de Hausdorff, 13
 - de Hilbert, 74
 - vectorial
 - normado, 15
 - topológico, 10
- estado del sistema, 44
- fórmula de Green, 40
- función
 - coerciva, 18
 - continua, 21
 - convexa, 9
 - cuadrática, 69
 - cuasiconvexa, 9
 - diferenciable, 21
 - estrictamente convexa, 10
 - fuertemente convexa, 48
 - indicatriz, 7, 36
 - inf-compacta, 14
 - inferiormente semicontinua, 13
 - Lipschitz-continua, 21
 - objetivo, 1
 - par, 49
 - propia, 7
 - subdiferenciable, 23
- funcional, 15

- identidad
 - de polarización, 75
 - del paralelogramo, 75
- inclusión diferencial, 43
- iteración, 44
- lema
 - de descenso, 64
 - de Opial, 48
- método
 - del máximo descenso, 51
 - directo, 19
- minimizador, 1
 - global, 11
 - local, 11
- norma, 74
- operador de proximidad, 36
- paso de discretización, 45
- producto interno, 73
- proyección ortogonal, 36
- punto crítico, 44
- reales extendidos, 7
- regla de Fermat, 2, 22
 - generalizada, 24
- resolvente, 36
- solución, 51
- subdiferencial, 23
- subgradiente, 23
- sucesión minimizante, 18
- teorema
 - de Hahn-Banach, 16
 - de Kakutani-Eberlein-Šmulian, 17
 - de Moreau-Rockafellar, 31
 - de Weierstrass, 2, 14
 - de Weierstrass-Hilbert-Tonelli, 14
- topología
 - débil, 15
 - fuerte, 15
- trayectoria, 44
- valor, 1

Bibliografía

- [1] F. Álvarez & J. Peypouquet: Asymptotic almost-equivalence of Lipschitz evolution systems in Banach spaces. *Nonlinear Anal., Theory Methods Appl.*, **73** (2010), 3018–3033.
- [2] D. Bertsekas: Nonlinear programming, Athena Scientific, 1999.
- [3] J. Borwein & A. Lewis: Convex analysis and nonlinear optimization, Springer, 2010.
- [4] H. Brézis: Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert, North Holland Publishing Company, 1973.
- [5] H. Brézis: Functional analysis, sobolev spaces and partial differential equations, Springer, 2011.
- [6] M.-G. Crandall & T.-M. Liggett: Generation of semigroups of nonlinear transformations on general Banach spaces, *Am. J. Math.*, **93** (1971), 265–298.
- [7] I. Ekeland & R. Temam: Analyse convexe et problèmes variationnels, Dunod, 1974.
- [8] Y. Kobayashi: Difference approximation of Cauchy problems for quasi-dissipative operators and generation of nonlinear semigroups, *J. Math. Soc. Japan*, **27** (1975), 640–665.
- [9] J.-J. Moreau, Propriétés des applications “prox”, *CRAS*, **256** (1963), 1069–1071.

- [10] J. Peypouquet, Asymptotic convergence to the optimal value of diagonal proximal iterations in convex minimization, *Journal Convex Anal.*, **16** (2009), 277–286.
- [11] J. Peypouquet & S. Sorin: Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time, *J. Convex Anal.* **17** (2010), 1113–1163.
- [12] R.-T. Rockafellar: Convex analysis, Princeton University Press, 1972.

Asociación Matemática Venezolana

Presidente: Rafael Sánchez Lamonedá

Consejo Directivo Nacional

Rafael Sánchez Lamonedá
Capítulo Capital

Alexander Carrasco
Capítulo de Centro Occidente

Oswaldo Araujo
Capítulo de Los Andes

Said Kas-Danouche
Capítulo de Oriente

Oswaldo Larreal
Capítulo Zuliano

La Asociación Matemática Venezolana fue fundada en 1990 como una organización civil sin fines de lucro cuya finalidad es trabajar por el desarrollo de las matemáticas en Venezuela.

Asociación Matemática Venezolana
Apartado 47.898, Caracas 1041-A, Venezuela
<http://amv.ivic.gob.ve>

Instituto Venezolano de Investigaciones Científicas

Consejo Directivo

Director

Eloy Sira

Subdirector

Alexander Briceño

**Representantes del Ministerio del Poder Popular para la
Ciencia, Tecnología e Innovación**

Guillermo Barreto

Juan Luis Cabrera

**Representante del Ministerio del Poder Popular para la
Educación Universitaria**

Prudencio Chacón

Representantes Laborales

José Garzaro

Víctor Peña

William Espinoza (Suplente)

Sirvia Ávila (Suplente)

Gerencia General

Lira Parra

Comisión Editorial

Eloy Sira (Coordinador)

Lucía Antillano

Horacio Biord

Jesús Eloy Conde

María Teresa Curcio

Rafael Gassón

Pamela Navarro

Héctor Suárez

Erika Wagner



Gobierno **Bolivariano**
de Venezuela

Ministerio del Poder Popular
para **Ciencia, Tecnología e Innovación**

