UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

APUNTE MAT279

curso obligatorio de la carrera
INGENIERÍA CIVIL MATEMÁTICA

OPTIMIZACIÓN NO LINEAL.

3ra Edición

Luis BRICEÑO • **Cristopher HERMOSILLA**

Departamento de Matemática
Julio 2023

Prefacio

Este apunte ha sido redactado con la finalidad de proveer a estudiantes de los programas de Ingeniería de la Universidad Técnica Federico Santa María con herramientas básicas de *Optimización No Lineal*[1]. Estas notas cubren aspectos generales de la optimización en espacio abstracto, así como resultados más específicos para espacios de Hilbert. Los contenidos cubren resultados de existencia de soluciones, caracterizaciones de estas, criterios analíticos para encontrarlas (condiciones de optimalidad) y también métodos iterativos para aproximar soluciones óptimas.

Las notas aquí presentadas fueron organizadas de forma tal de cubrir los contenidos del curso *Optimización No Lineal (MAT279)* que imparte regularmente el Departamento de Matemática de la Universidad Técnica Federico Santa María. Este curso es parte de la malla de la carrera *Ingeniería Civil Matemática*, y como tal requiere herramientas abstractas de *Análisis*. Sin embargo, todos los resultados expuestos en el apunte han sido escritos de forma general, por lo cual cualquier estudiante de ingeniería con un conocimiento básico en *Análisis en* $\mathbb{R}^n$ y *Álgebra Lineal* puede comprender el material expuesto en estas notas.

Esta es la tercera versión del apunte, y pese a que muchos errores tipográficos fueron corregidos, aún pueden quedar algunos. Todo posible error que el lector pueda encontrar en las notas es de nuestra exclusiva responsabilidad. Agradecemos hacer llegar comentarios y observaciones a cualquiera de los autores.

Luis BRICEÑO
Campus San Joaquín
Santiago

•

Cristopher HERMOSILLA
Casa Central
Valparaíso

[1] El término *No Lineal* debe ser entendido en este contexto como *No Necesariamente Lineal*.

Notación básica

Conjuntos básicos

$\mathbb{R}$	Números Reales.
$\mathbb{R}^n$	Conjunto de n-tuplas de Números Reales.
$\mathbb{R} \cup \{+\infty\}$	Números Reales (superiormente) extendidos.
$\mathbb{N}$	Números Naturales
$\mathbb{M}_{n\times m}(\mathbb{R})$	Matrices a coeficientes reales de dimensión $n \times m$
$\mathbb{S}^n$	Matrices reales simétricas de dimensión n
$\mathbb{S}^n_+(\mathbb{R})$	Matrices reales simétricas semi-definidas positivas de dimensión n
$\mathbb{S}^n_{++}(\mathbb{R})$	Matrices reales simétricas definidas positivas de dimensión n

Conjuntos Genéricos

$\mathbf{X}$	Espacio ambiente
$\mathbf{S}$	Conjunto de restricciones
$\mathbb{B}_{\mathbf{X}}(x,r)$	Bola cerrada de radio $r > 0$ y centro $x \in \mathbf{X}$ de un espacio métrico $(\mathbf{X}, d)$
$\mathbb{B}_{\mathbf{X}}$	Bola cerrada unitaria de un espacio vectorial normado $(\mathbf{X}, \|\cdot\|)$
$\operatorname{int}\mathbf{S}$	interior de $\mathbf{S}$
$\overline{\mathbf{S}}$	adherencia de $\mathbf{S}$

Conjuntos Especiales

$\operatorname{dom}(f)$	Dominio efectivo de $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$
$\operatorname{epi}(f)$	Epígrafo de $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$
$\Gamma_\gamma(f)$	Conjunto de subnivel de $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ y parámetro $\gamma \in \mathbb{R}$
$\arg\min_{\mathbf{X}}(f)$	Conjunto de mínimos de $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$

Normas y productos internos

$\lvert x\rvert = \sqrt{\sum_{i=1}^{n} x_i^2}$	Norma Euclideana de $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$
$\|\cdot\|$	Norma de un espacio vectorial arbitrario $\mathbf{X}$
$x^\top y = \sum_{i=1}^{n} x_i y_i$	Producto interno de $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ e $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$
$\langle \cdot, \cdot \rangle$	Producto interno de un espacio Euclideano arbitrario $\mathbf{X}$

Operadores funcionales

∇f	Gradiente de $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$
$Df(x)(\cdot)$	Diferencial de Gâteaux de $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ en $x \in \mathbf{X}$
$\nabla^2 f$	Matriz Hessiana de $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$
$D^2 f(x)(\cdot, \cdot)$	Segundo Diferencial de Gâteaux de $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ en $x \in \mathbf{X}$

Índice general

CAPÍTULO 1

Introducción a la Optimización

El objetivo central de este curso es estudiar problemas de optimización:

(P) Minimizar $f(x)$ sobre todos los $x \in \mathbf{X}$ que satisfacen la restricción $x \in \mathbf{S}$.

En nuestro contexto $\mathbf{X}$ será un espacio vectorial, la función $f : \mathbf{X} \to \mathbb{R}$ será un criterio a minimizar, el cual llamaremos *función objetivo* y el conjunto $\mathbf{S} \subseteq \mathbf{X}$ representará las *restricciones* impuestas sobre el problema de interés.

El valor numérico que toma el problema (P) está dado por

$$\operatorname{val}(\mathrm{P}) = \operatorname{ínf}_{\mathbf{S}}(f) := \inf_{x \in \mathbf{X}} \{f(x) \mid x \in \mathbf{S}\},$$

el cual queda bien definido si adoptamos la convención $\operatorname{val}(\mathrm{P}) = +\infty := \sup\{\mathbb{R}\}$ para el caso $\mathbf{S} = \emptyset$. Por otra parte, una solución del problema (P) será llamada *óptimo* o *mínimo*, y corresponderá a un punto $\bar{x} \in \mathbf{X}$ que verifique la condición

$$f(\bar{x}) \leq f(x) \text{ para todo } x \in \mathbf{X} \text{ tal que } x \in \mathbf{S}.$$

Una solución óptima del problema (P) se dirá *estricto* si la condición anterior se tiene con desigualdad estricta para todos los puntos diferentes al mínimo, es decir

$$f(\bar{x}) < f(x) \text{ para todo } x \in \mathbf{X} \text{ tal que } x \in \mathbf{S} \setminus \{\bar{x}\}.$$

En caso de haber un óptimo, y para para enfatizar la existencia de éste, el valor numérico que toma el problema (P) se escribirá

$$\operatorname{val}(\mathrm{P}) = \operatorname{mín}_{\mathbf{S}}(f) := \min_{x \in \mathbf{X}} \{f(x) \mid x \in \mathbf{S}\}.$$

El conjunto de soluciones del problema (P) se denotará por

$$\operatorname{sol}(\mathrm{P}) = \arg\operatorname{mín}_{\mathbf{S}}(f) := \{x \in \mathbf{S} \mid f(x) = \operatorname{val}(\mathrm{P})\}.$$

En este capítulo nos enfocaremos en la existencia de mínimos para el problema (P) en un contexto abstracto, es decir, en criterios para determinar que el conjunto $\operatorname{sol}(\mathrm{P})$ sea no vacío. En particular, estudiaremos la noción de semicontinuidad inferior y algunas nociones de compacidad.

Observación 1.1. *Notemos que* $\sup_{\mathbf{S}}(f) := \sup_{x \in \mathbf{X}} \{f(x) \mid x \in \mathbf{S}\} = -\operatorname{ínf}_{\mathbf{S}}(-f)$. *Por lo tanto la teoría que desarrollaremos en este curso puede ser igualmente aplicada a problemas donde se busca maximizar la función objetivo en vez de minimizarla, tomando en cuenta el cambio de signo descrito anteriormente. Formulaciones del tipo maximización aparecen típicamente en Economía.*

1.1. Clases de problemas de optimización destacados

Antes de continuar con la teoría, revisaremos algunos problemas de optimización cuyas estructuras los hacen fácilmente reconocibles.

1.1.1. Programación lineal

Esta clase de problemas busca minimizar una función objetivo lineal sobre el espacio $\mathbf{X} = \mathbb{R}^n$

$$f(x) = c^\top x = \sum_{i=1}^{n} c_i x_i$$

donde $c \in \mathbb{R}^n$, y sujeto a un conjunto de restricciones que se pueden escribir como poliedros

$$\mathbf{S} = \{x \in \mathbb{R}^n \mid \ Ax \leq b,\ x \geq 0\},$$

con $A \in \mathbb{M}_{n\times m}(\mathbb{R})$, una matriz a coeficientes reales de dimensión $n \times m$, y $b \in \mathbb{R}^m$.

Problemas de este estilo aparecen frecuentemente en economía, donde la función objetivo representa un costo o bien una utilidad (visto como problema de maximización).

1.1.2. Programación semidefinida

Esta clase de problemas es el análogo de la programación lineal sobre el espacio vectorial de matrices simétricas de dimensión n, que denotamos por $\mathbb{S}^n(\mathbb{R})$. Se busca minimizar una función objetivo lineal

$$f(X) = \operatorname{tr}(CX) = \sum_{i,j=1}^{n} C_{ij} X_{ij}$$

con $C \in \mathbb{S}^n(\mathbb{R})$ sujeto a un conjunto de restricciones que se pueden escribir como

$$\mathbf{S} = \{X \in \mathbb{S}^n(\mathbb{R}) \mid \ \operatorname{tr}(A_i X) = b_i,\ i = 1,\ldots,m,\ X \succeq 0\},$$

con $A_1,\ldots,A_m \in \mathbb{S}^n(\mathbb{R})$, matrices dadas y $b_1,\ldots,b_m \in \mathbb{R}$. La notación $X \succeq 0$ para $X \in \mathbb{S}^n(\mathbb{R})$ significa que X es semi-definida positiva.

1.1.3. Optimización Espectral

Muchas veces, cuando se trabaja con matrices, es más importante conocer sus valores propios que la matriz misma. La optimización espectral corresponde a problemas donde la función objetivo depende de los valores propios de una matriz y no directamente de la matriz. Al igual que en el caso anterior, el problema se plantea sobre el espacio $\mathbb{S}^n(\mathbb{R})$. Recordemos que si $X \in \mathbb{S}^n(\mathbb{R})$, entonces sus n valores propios son Reales. Esto permite definir la función espectral $\lambda : \mathbb{S}^n(\mathbb{R}) \to \mathbb{R}^n$ por

$$\lambda(X) = (\lambda_1(X),\ldots,\lambda_n(X))$$

donde $\lambda_1(X) \geq \ldots \geq \lambda_n(X)$ son valores propios de X ordenados de forma decreciente. Luego un problema de optimización espectral corresponde a minimizar una función objetivo del tipo

$$f(X) = g \circ \lambda(X) = g(\lambda(X)) = g(\lambda_1(X),\ldots,\lambda_n(X))$$

con $g : \mathbb{R}^n \to \mathbb{R}$ alguna función dada.

1.1.4. Control Óptimo en tiempo discreto

Esta clase de problemas, el primero en dimensión infinita que mencionaremos, consiste en minimizar un funcional cuyo argumento es una sucesión generada por una regla de recurrencia inductiva

$$x_{k+1} = \phi(x_k, u_k), \quad \forall k \in \mathbb{N},$$

donde $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ es un campo vectorial dado. El problema consiste en minimizar, para un cierto costo $g : \mathbb{R}^n \times \mathbb{R}^m$ y factor de descuento $\lambda \geq 0$, un funcional del tipo

$$f(\{x_k\}, \{u_k\}) = \sum_{k=1}^{\infty} e^{-\lambda k} g(x_k, u_k)$$

En este caso, los espacio naturales para estudiar el problema son $\mathbf{X} = \ell_p(\mathbb{R}^n) \times \ell_q(\mathbb{R}^m)$, donde $\ell_r(\mathbb{R}^N)$ es el espacio de la sucesiones $\{a_k\}$ en $\mathbb{R}^N$ tales que la siguiente serie converge

$$\sum_{k=0}^{\infty} |a_k|^r < +\infty.$$

1.1.5. Cálculo de Variaciones

Esta clase de problemas, también de dimensión infinita, consiste en minimizar un funcional cuyo argumento es una curva en el espacio $x : [a,b] \to \mathbb{R}^n$:

$$f(x) = \int_a^b L(t, x(t), \dot{x}(t)) dt,$$

donde $L : [a,b] \times \mathbb{R}^n \times \mathbb{R}^n$ es una función llamada *Lagrangiano*. El espacio natural para plantear tales problemas es $\mathbf{X} = AC^n[a,b]$, el espacio espacio de la curvas absolutamente continuas, es decir las funciones $x : [a,b] \to \mathbb{R}^n$ que satisfacen, para ciertos $\xi_1, \ldots, \xi_n \in \mathbb{R}$ y $v_1, \ldots, v_n \in L^1[a,b]$

$$x_i(t) = \xi_i + \int_a^t v_i(s) ds, \qquad \forall t \in [a,b],\ \forall i = 1, \ldots, n.$$

Problemas de Cálculo de Variaciones, típicamente buscan minimizar el costo integral descrito anteriormente, sujeto a que los puntos extremos de las trayectorias están previamente prescritos, es decir, para ciertos $\alpha, \beta \in \mathbb{R}^n$, el conjunto de restricciones está dado por

$$\mathbf{S} = \{x \in AC^n[a,b] \mid\ x(a) = \alpha,\ x(b) = \beta\}.$$

Esta clase de problemas de optimización aparecen muchas veces en mecánica, donde x representa la trayectoria de una partícula en el espacio y $\dot{x}$ su velocidad.

Mencionamos también una clase particular de problemas de Cálculo de Variaciones, donde además de la restricción sobre los puntos extremos de la trayectoria, se considera una restricción integral del estilo

$$\int_a^b g(t, x(t), \dot{x}(t)) = c.$$

Estos problemas se conocen como problemas *isoperimétricos* y su nombre está motivado por problemas en el plano $\mathbb{R}^2$ donde el largo de la curva está fijo, es decir:

$$\int_a^b \sqrt{(\dot{x}_1(t))^2 + (\dot{x}_2(t))^2} dt = c.$$

1.1.6. Control Óptimo en tiempo continuo

Esta clase de problemas son una extensión de los problemas de Cálculo de Variaciones y corresponden a problemas donde la velocidad de las trayectorias están determinada por una ecuación diferencial ordinaria que dependen de un parámetro (el control). Más aún, el funcional a ser minimizado puede considerar costos explícitos sobre los puntos extremos, es decir, en Control Óptimo se busca minimizar un funcional del tipo

$$\int_a^b \mathcal{L}(t,x(t),u(t))dt + g(x(a),x(b))$$

sujeto a una restricción dinámica sobre la velocidad

$$\dot{x}(t) = \phi(t,x(t),u(t)), \quad \text{y} \quad u(t) \in U \subseteq \mathbb{R}^m \quad \text{c.t.p } t \in [a,b]$$

donde $u : [a,b] \to \mathbb{R}^m$ es una función medible, llamada *control* o *input*. En este caso tenemos que $\mathcal{L} : [a,b] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ es una función de costo acumulativa mientras que $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ es una función de costo sobre los puntos extremos de la trayectoria.

Bajo condiciones estándar, la ecuación diferencial está bien puesta, en el sentido que cada función medible $u : [a,b] \to \mathbb{R}^m$ produce una única solución si la condición inicial está dada. Lo que implica, en principio, que el espacio natural para buscar mínimos es el conjunto de funciones medibles valores en el conjunto $U \subseteq \mathbb{R}^m$. Este espacio tiene pocas propiedades topológicas favorables, lo cual no lo hace un buen candidato para plantear los problemas de control. En cursos más avanzados se verá que tal dificultad puede ser salvada usando teoremas de selección y una formulación equivalente sobre el espacio $\mathbf{X} = AC^n[a,b]$.

1.2. Problemas industriales de actualidad

Ahora mencionaremos algunos problema de optimización que son actualmente utilizados en aplicaciones industriales de interés práctico. Estos problemas serán en particular nuestra principal motivación para estudiar métodos numéricos en capítulos más avanzados.

1.2.1. Compresión y recuperación de imágenes

Consideremos una imagen de $n \times m$ pixeles (con $N = nm$ grande) en escala de grises representada por una matriz $\bar{X} \in \mathbb{M}_{n\times m}([0,255])$, donde para todo $i \in \{1,\dots,n\}$ y $j \in \{1,\dots,m\}$, la componente (i,j) de la matriz $\bar{X}$, denotada $\bar{X}_{ij}$ representa la intensidad de luminosidad del pixel (i,j), que puede variar entre 0 (negro) y 255 (blanco).

La imagen se quiere comprimir a través de una matriz conocida $A \in \mathbb{M}_{p\times N}(\mathbb{R})$ de modo que $z := A\bar{x} \in \mathbb{R}^p$ es la imagen comprimida ($p \ll N$), donde $\bar{x} \in \mathbb{R}^N$ es un vector que representa a la matriz $\bar{X}$ via la relación

$$\bar{x}_{n(j-1)+i} = \bar{X}_{ij}, \quad \forall i \in \{1,\dots,n\},\ j \in \{1,\dots,m\}.$$

Luego el problema de recuperación de imágenes consiste en encontrar una buena aproximación de $\bar{x}$, conociendo z, bajo supuestos *a priori* sobre $\bar{x}$.

Se dice que la imagen original es parsimoniosa (*sparse* en inglés) en alguna base ortonormal $v_1,\dots,v_N$ (llamada *wavelet*), si $\bar{x}=\sum_{i=1}^N y_i v_i$, pocos y_i no nulos. Muchas imágenes son parsimoniosas en algunas bases de wavelets, lo que indica que la imagen puede ser muy bien representada a través de pocos elementos de la base. Notar que si $F\in\mathbb{M}_{N\times N}(\mathbb{R})$ es la matriz cuadrada que tiene como columnas los vectores ortonormales $v_1,\dots,v_N$, se tiene que $\bar{x}=Fy$, con $y=(y_1,\dots,y_N)$ y $F^\top=F^{-1}$, de donde $y=F^\top\bar{x}$.

Si suponemos que la imagen $\bar{x}$ es parsimoniosa con respecto a $v_1,\dots,v_N$, entonces el vector $y=F^\top\bar{x}$ tiene muchas componentes nulas lo que significa que

$$\|y\|_0 := |\{i\in\{1,\dots,N\} : y_i\neq 0\}|$$

es un número pequeño. Por lo tanto, una manera de aproximar $\bar{x}$ es considerar el problema

$$\text{Minimizar } \|F^\top x\|_0 \text{ sobre todos los } x\in\mathbb{R}^N \text{ que satisfacen la restricción } Ax=z.$$

Como la función $x\mapsto\|x\|_0$ tiene malas propiedades, una relajación ampliamente usada en restauración de imágenes consiste en usar la norma $\|y\|_1=\sum_{i=1}^N|y_i|$, de donde se obtiene el problema

$$\text{Minimizar } \|F^\top x\|_1 \text{ sobre todos los } x\in\mathbb{R}^N \text{ que satisfacen la restricción } Ax=z.$$

A su vez, una forma de aproximar el problema anterior es usar penalización del tipo

$$\text{Minimizar } \|F^\top x\|_1+\frac{1}{\lambda}|Ax-z|^2 \text{ sobre todos los } x\in\mathbb{R}^N.$$

donde $\lambda>0$ es un parámetro que modela cuánta preferencia al ajuste $Ax=z$ se tiene sobre la parsimonia de $F^\top x$. La ventaja de este último problema es que no tiene restricciones adicionales.

1.2.2. Mercado de uso de suelo

Consideremos una ciudad con n zonas y m tipos de hogares que buscan localizarse, indexados por $i\in\{1,\dots,n\}$ y $h\in\{1,\dots,m\}$, respectivamente. Para cada zona $i\in\{1,\dots,n\}$ y tipo de hogar $h\in\{1,\dots,m\}$, denotamos S_i la oferta inmobiliaria en la zona i y H_h el número de hogares de tipo h que buscan localizarse. Por simplicidad, supondremos que el mercado está en equilibrio, es decir, que hay tantas casas disponibles como hogares a localizarse en la ciudad. Esto se representa en términos matemáticos como sigue:

$$\sum_{i=1}^n S_i=\sum_{h=1}^m H_h,$$

Por otra parte, supondremos que se conocen las preferencias de cada tipo de hogar en cada zona. Más precisamente, tenemos acceso a C_{hi} que es una medida de utilidad percibida por un hogar tipo $h\in\{1,\dots,m\}$ en la zona $i\in\{1,\dots,n\}$. En este problema se busca una localización de hogares en zonas tal que se maximice la utilidad total de los hogares y se satisfagan las restricciones de oferta y demanda. Más precisamente, el problema es

$$\begin{cases}\text{Maximizar} & \displaystyle\sum_{i=1}^n\sum_{h=1}^m C_{hi}X_{hi} \text{ sobre los } X\in\mathbb{M}_{m\times n}(\mathbb{R})\\ \text{tales que} & \sum_{i=1}^n X_{hi}=H_h,\ \forall h=1,\dots,m\\ & \sum_{h=1}^m X_{hi}=S_i,\ \forall i=1,\dots,n\\ & X_{ij}\geq 0,\ \forall i=1,\dots,n,\ ,\ \forall h=1,\dots,m.\end{cases}$$

La componente X_{hi} de la matriz $X \in \mathbb{M}_{m\times n}(\mathbb{R})$ representa en este caso la cantidad de hogares tipo h que se localizan en la zona i. Este es problema se puede formular como un problema de programación lineal y puede ser resuelto por el método simplex. Las soluciones de este tipo de problemas se encuentran en los extremos del poliedro que generan las restricciones lineales y son altamente sensibles a los valores de las utilidades de la matriz C, pudiendo pasar X_{hi} de 0 a H_h si C_{hi} pasa de no ser el máximo valor entre $C_{h1},\dots,C_{hN}$ a serlo, por ejemplo. En el caso en que existe incertidumbre en la estimación de las utilidades, en la literatura es ampliamente utilizado agregar una penalización entrópica, obteniendo el problema

$$\begin{cases} \text{Maximizar} & \displaystyle\sum_{i=1}^{n}\sum_{h=1}^{m} C_{hi}X_{hi} + \frac{1}{\lambda}X_{hi}(\log(X_{hi})-1) \text{ sobre los } X \in \mathbb{M}_{m\times n}(\mathbb{R}) \\ \text{tales que} & \sum_{i=1}^{n} X_{hi} = H_h,\ \forall h = 1,\dots,m \\ & \sum_{h=1}^{m} X_{hi} = S_i,\ \forall i = 1,\dots,n \end{cases}$$

La función $X \mapsto -\sum_{i=1}^{n}\sum_{h=1}^{m} X_{hi}(\log(X_{hi})-1)$ está muy relacionada con la entropía de Shannon que mide el nivel de incertidumbre de variables aleatorias. Esta modificación permite evitar grandes cambios de la solución a modificaciones menores de las variables C_{hi}. Este problema será objeto de estudio en este curso.

1.3. Funciones a valores en $\mathbb{R}\cup\{+\infty\}$

En el análisis que llevaremos a cabo en la primera parte del curso será conveniente considerar funciones cuyos valor pertenecen a la *recta Real (superiormente) extendida* $\mathbb{R}\cup\{+\infty\} = (-\infty,+\infty]$ y no solamente en $\mathbb{R} = (-\infty,+\infty)$. La principal ventaja de hacer esto se describe a continuación:

Definamos $\delta_{\mathbf{S}} : \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$, la *función indicatriz* del conjunto $\mathbf{S}$, via

$$\delta_{\mathbf{S}}(x) := \begin{cases} 0 & x \in \mathbf{S}, \\ +\infty & x \notin \mathbf{S}. \end{cases}$$

Usando la convención

$$\alpha + (+\infty) = (+\infty) + \alpha = +\infty, \quad \forall \alpha \in \mathbb{R}$$

tenemos que

$$\text{val}\,(\text{P}) = \inf_{x\in\mathbf{X}}\{f(x) + \delta_{\mathbf{S}}(x)\}.$$

De esta manera, el problema (P) se puede formular como un problema sin restricciones, pero con una función objetivo a valores en la recta Real extendida. Esto permite tratar problemas de optimización abstracta de una forma unificada, independiente del conjunto de restricciones $\mathbf{S}$, cuya información estará incluida implícitamente en la función objetivo.

1.3.1. Definiciones básicas

El estudio de problemas de optimización con funciones objetivo a valores en la recta Real extendida debe ser manejado con cuidado. En particular, nuevas definiciones y convenciones tienen

que ser introducidas. Por ejemplo, dada una función $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$, su dominio efectivo es el conjunto

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

Además, diremos que $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ es propia si $\mathrm{dom}(f) \neq \emptyset$.

En lo que sigue, y a menos que se diga otra cosa, asumiremos que la función objetivo tiene valores sobre la recta Real extendida, es decir, $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$. Además, obviaremos la presencia de restricciones, las cuales asumiremos se encuentran implícitamente definidas en la estructura de la función objetivo via la relación

$$\mathbf{S} = \mathrm{dom}(f).$$

Bajo estas circunstancias tendremos que

$$\text{ínf}_{\mathbf{X}}(f) := \mathrm{val}\,(\mathrm{P}) = \inf_{x \in \mathbf{X}} \{f(x) \mid \ x \in \mathrm{dom}(f)\} \quad \text{y} \quad \arg\min_{\mathbf{X}}(f) := \mathrm{sol}\,(\mathrm{P}).$$

El *conjunto de nivel inferior (o subnivel) de parámetro* $\gamma \in \mathbb{R}$ de una función $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ está dado por

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

y su *epígrafo* es el subconjunto de $\mathbf{X} \times \mathbb{R}$ dado por

$$\mathrm{epi}(f) := \{(x, \lambda) \in \mathbf{X} \times \mathbb{R} \mid \ f(x) \leq \lambda\}.$$

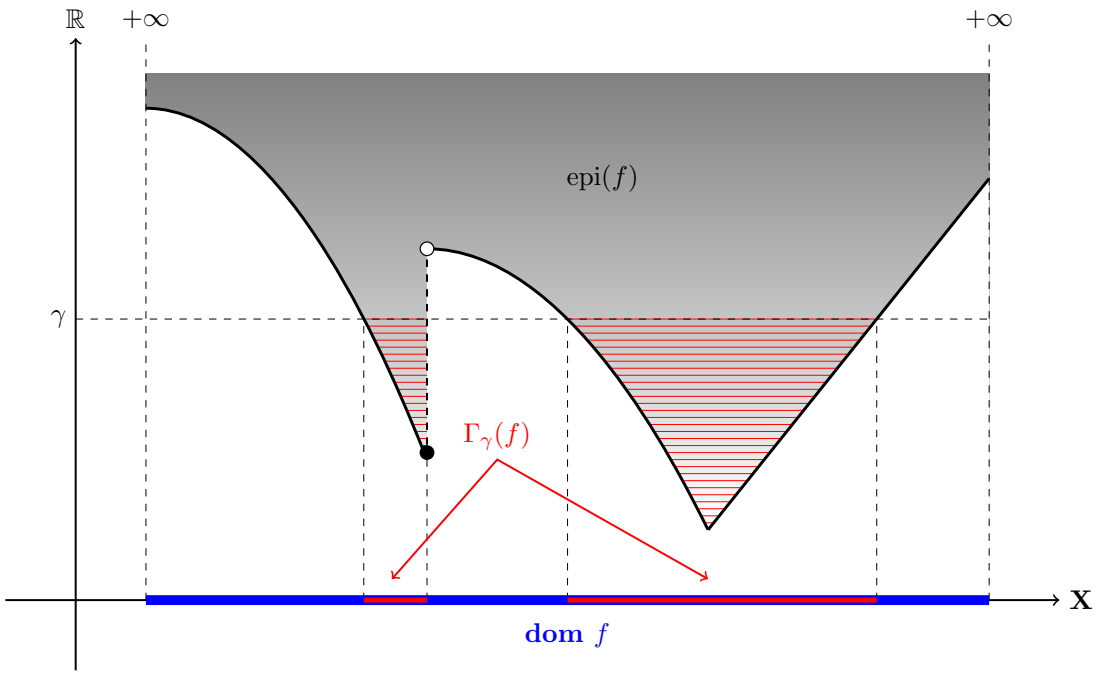


Figura 1.1: Subniveles y epígrafo de una función

1.3.2. Convenciones algebraicas

Dados $\alpha > 0$ y funciones $f, g : \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$, para darle sentido a la expresión $f + \alpha g$ introducimos la siguientes reglas algebraicas en $\mathbb{R}\cup\{+\infty\}$ que generalizan las conocidas en $\mathbb{R}$:

1. $(+\infty) + \alpha = \alpha + (+\infty) = +\infty$, para todo $\alpha \in \mathbb{R}\cup\{+\infty\}$.
2. $\alpha \cdot (+\infty) = (+\infty) \cdot \alpha = +\infty$, para todo $\alpha > 0$.
3. $0 \cdot (+\infty) = (+\infty) \cdot 0 = 0$.

Observación 1.2. *Bajo estas condiciones el producto no es continuo en el sentido que si* $\alpha_k \to \alpha$ *and* $\beta_k \to \beta$*, con* $\alpha, \beta \in \mathbb{R}\cup\{+\infty\}$*, uno no tiene necesariamente que* $\alpha_k\beta_k \to \alpha\beta$.

CAPÍTULO 2

Existencia de mínimos

Hasta el momento no hemos necesitado mayor estructura sobre el espacio $\mathbf{X}$, pero para asegurar la existencia de mínimos y para el resto del curso trabajaremos en el contexto de espacios de Hilbert $(\mathbf{X}, \langle \cdot, \cdot \rangle)$. Recordemos la definición y algunas propiedades de los espacios de Hilbert estudiadas en MAT225: Análisis I.

2.1. Espacios de Hilbert

Sea $\mathbf{X}$ es un espacio vectorial. Un producto interno para $\mathbf{X}$ es una función $\langle \cdot, \cdot \rangle \colon \mathbf{X} \times \mathbf{X} \to \mathbb{R}$ bilineal, simétrica y que satisface, para todo $x \neq 0$, $\langle x, x \rangle > 0$. Bajo estas condiciones, la función $\| \cdot \| \colon \mathbf{X} \to [0, +\infty[\colon x \mapsto \sqrt{\langle x, x \rangle}$ define una norma para $\mathbf{X}$. Si $(\mathbf{X}, \| \cdot \|)$ es un espacio vectorial normado completo (Banach), se dice que $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ es un espacio de Hilbert.

Lema 2.1 (Teorema de Representación de Riesz)**.** *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbert y sea* $\mathbf{X}^* = \{\ell \colon \mathbf{X} \to \mathbb{R} : \ell \text{ es lineal y continua}\}$ *el dual topológico de* $\mathbf{X}$*. Entonces*

$$(\forall \ell \in \mathbf{X}^*)(\exists z_\ell \in \mathbf{X}) \quad \ell = \langle z_\ell, \cdot \rangle.$$

El teorema de representación de Riesz permite identificar a $\mathbf{X}^*$ con $\mathbf{X}$ y hace a $\mathbf{X}$ un espacio reflexivo.

En un espacio de Hilbert $(\mathbf{X}, \langle \cdot, \cdot \rangle)$, aparte de la topología de la norma o topología fuerte (denotada $\mathcal{T}_{\|\cdot\|}$), es común considerar la topología débil, denotada $\sigma(\mathbf{X}, \mathbf{X}^*)$ (también denotada $\sigma(\mathbf{X}, \mathbf{X})$ gracias a la identificación producto del Teo. Riesz). Ésta es la topología menos fina (con menos abiertos) que mantiene la continuidad de la familia de funciones en $\mathbf{X}^*$. De su definición, es claro que $\sigma(\mathbf{X}, \mathbf{X}) \subset \mathcal{T}_{\|\cdot\|}$ y se puede probar que $A \in \sigma(\mathbf{X}, \mathbf{X})$ si y sólo si

$$\forall x \in A,\ \exists x_1, \ldots, x_n \in \mathbf{X},\ \exists \varepsilon > 0,\ \{y \in \mathbf{X} \mid |\langle x_i, y - x \rangle| < \varepsilon,\ \forall i = 1, \ldots, n\} \subseteq A.$$

Recuerdo : Convergencia en espacios de Hilbert

Dada $\mathcal{T} \in \{\mathcal{T}_{\|\cdot\|}, \sigma(\mathbf{X},\mathbf{X})\}$, una sucesión $\{x_n\}_{n\in\mathbb{N}}$ y x en $\mathbf{X}$, se define

$$x_n \to x \text{ en } \mathcal{T} \quad \Leftrightarrow \quad (\forall U \in \mathcal{N}_x)(\exists n_0 \in \mathbb{N})(\forall n \geq n_0) \quad x_n \in U,$$

donde $U \in \mathcal{N}_x$ si y sólo si $x \in U \in \mathcal{T}$.
En particular, si $\mathcal{T} = \mathcal{T}_{\|\cdot\|}$, $x_n \to x$ en $\mathcal{T}$ se denota simplemente como $x_n \to x$, se dice que $\{x_n\}_{n\in\mathbb{N}}$ *converge fuertemente a* x y se tiene

$$x_n \to x \quad \Leftrightarrow \quad \|x_n - x\| \to 0.$$

Por otra parte, si $\mathcal{T} = \sigma(\mathbf{X},\mathbf{X})$, $x_n \to x$ en $\mathcal{T}$ se denota como $x_n \rightharpoonup x$, se dice que $\{x_n\}_{n\in\mathbb{N}}$ *converge débilmente a* x y se tiene

$$x_n \rightharpoonup x \quad \Leftrightarrow \quad (\forall y \in \mathbf{X}) \quad \langle y, x_n \rangle \to \langle y, x \rangle.$$

Recordamos que un conjunto C es cerrado para $\mathcal{T} \in \{\mathcal{T}_{\|\cdot\|}, \sigma(\mathbf{X},\mathbf{X})\}$ si y solo si $X \setminus C = C^c \in \mathcal{T}$ (es abierto). Mientras que la topología $\mathcal{T}_{\|\cdot\|}$ es metrizable (considerando $d\colon (x,y) \mapsto \|x-y\|$), la topología débil $\sigma(\mathbf{X},\mathbf{X})$ no es metrizable, lo que dificulta la caracterización de cerradura via sucesiones.

Recuerdo : Cerradura en espacios de Hilbert

Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $\mathcal{T} \in \{\mathcal{T}_{\|\cdot\|}, \sigma(\mathbf{X},\mathbf{X})\}$. En general, se tiene que si $C \subset \mathbf{X}$ es cerrado en $\mathcal{T}$ entonces

$$(\forall \{x_n\}_{n\in\mathbb{N}} \subset C) \quad [x_n \to x \text{ en } \mathcal{T} \quad \Rightarrow \quad x \in C].$$

En particular, si $\mathcal{T} = \mathcal{T}_{\|\cdot\|}$, en vez de C cerrado en $\mathcal{T}$ decimos que C es cerrado fuerte y si $\mathcal{T} = \sigma(\mathbf{X},\mathbf{X})$ decimos que C es cerrado débil. Como $\mathcal{T}_{\|\cdot\|}$ es metrizable, se tiene la equivalencia

$$C \text{ es cerrado fuerte} \quad \Leftrightarrow \quad (\forall \{x_n\}_{n\in\mathbb{N}} \subset C) \quad [x_n \to x \;\Rightarrow\; x \in C].$$

2.2. Semicontinuidad inferior

Las topologías definidas más arriba nos permiten definir la semicontinuidad inferior, que es el primer ingrediente para obtener existencia de mínimos.

Definición 2.1. *Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y sea $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función. Se dice que f es semicontinua inferior respecto a la topología $\mathcal{T} \in \{\sigma(\mathbf{X},\mathbf{X}), \mathcal{T}_{\|\cdot\|}\}$ (abreviado $\mathcal{T}$-s.c.i. o simplemente s.c.i. si la topología es clara del contexto) si y sólo si todos sus conjuntos de nivel inferior son cerrados para $\mathcal{T}$ o, equivalentemente,*

$$(\forall \gamma \in \mathbb{R}) \quad \mathbf{X} \setminus \Gamma_\gamma(f) \in \mathcal{T}.$$

Como recordamos más arriba, la topología débil tiene menos abiertos que la topología fuerte, i.e., $\sigma(\mathbf{X},\mathbf{X}) \subset \mathcal{T}_{\|\cdot\|}$. Por lo tanto, toda función $\sigma(\mathbf{X},\mathbf{X})$-s.c.i. es $\mathcal{T}_{\|\cdot\|}$-s.c.i. pero la recíproca no es cierta.

La semicontinuidad inferior se estudia en ciertos cursos usando un enfoque puntual, es decir, se define para cada punto; esto contrasta con Definición 2.1 que está escrita como propiedad global de la función. En particular, puede ser familiar al lector la siguiente definición para funciones definidas sobre los números reales: $f : \mathbb{R} \to \mathbb{R}$ es semicontinua inferior en $x \in \mathbb{R}$ si y sólo si

$$f(x) \leq \liminf_{y \to x} f(y) := \sup_{\varepsilon > 0} \inf_{y \in \mathbb{R}} \{ f(x) \mid \ y \in (x-\varepsilon, x+\varepsilon) \}.$$

Veremos ahora que este criterio, y otros más, son definiciones equivalentes para la semicontinuidad inferior de una función.

Proposición 2.1. *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert, sea $\mathcal{T} \in \{\mathcal{T}_{\|\cdot\|}, \sigma(\mathbf{X}, \mathbf{X})\}$ y sea $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función. Las siguientes afirmaciones son equivalentes:*

(i) *f es $\mathcal{T}$-s.c.i. .*

(ii) *$\forall \gamma \in \mathbb{R}$, $\{x \in \mathbf{X} \mid \ f(x) > \gamma\} \in \mathcal{T}$.*

(iii) *$\forall x \in \mathbf{X}$, $f(x) \leq \liminf_{y \to x} f(y) := \sup_{A \in \mathcal{N}_x} \inf_{y \in A} f(y)$*

(iv) *$\forall x \in \mathbf{X}$, $\forall \gamma < f(x)$, $\exists A_\gamma \in \mathcal{N}_x$ tal que $\forall y \in A_\gamma$ tenemos que $f(y) > \gamma$.*

(v) *$epi(f)$ es cerrado para la topología $\mathcal{T} \times \mathcal{T}_{\mathbb{R}}$, donde $\mathcal{T}_{\mathbb{R}}$ es la topología usual de $\mathbb{R}$.*

Demostración. La demostración se descompone en varias partes[1]:

- $(i) \iff (ii)$ Trivial, por definición.
- $(ii) \implies (iii)$ Sea $x \in \mathbf{X}$ y $\gamma \in (-\infty, f(x))$ tenemos que $A = \{y \in \mathbf{X} \mid \ f(y) > \gamma\} \in \mathcal{N}_x$ ya que $A \in \mathcal{T}$ por (ii) y $x \in A$. De este modo, $\gamma \leq \liminf_{y \to x} f(y)$. Como lo anterior es válido para todo $\gamma < f(x)$, hacemos $\gamma \to f(x)$ y concluimos el resultado.
- $(iii) \implies (iv)$ Sea $x \in \mathbf{X}$ y $\gamma \in (-\infty, f(x))$. Por (iii), tenemos que $\gamma < \sup_{A \in \mathcal{N}_x} \inf_{y \in A} f(y)$. Usando la definición del supremo tenemos que existe $A \in \mathcal{N}_x$ tal que $\gamma < \inf_{y \in A} f(y)$, de donde concluimos fácilmente.
- $(iv) \implies (v)$ Tomemos $(x, \lambda) \notin \text{epi}(f)$, lo que equivale a $\lambda < f(x)$. Consideremos $\gamma \in \mathbb{R}$ tal que $\lambda < \gamma < f(x)$. Luego (iv) implica la existencia de $A_\gamma \in \mathcal{N}_x$ tal que $\forall y \in A_\gamma, \ f(y) > \gamma$, de modo que $(y, \gamma) \notin \text{epi}(f)$. Se sigue que $A_\gamma \times (-\infty, \gamma)$ y $\text{epi}(f)$ son disjuntos, y como $A_\gamma \times (-\infty, \gamma)$ es un abierto para la topología $\mathcal{T} \times \mathcal{T}_{\mathbb{R}}$ que contiene al punto (x, λ), concluimos que $\mathbf{X} \setminus \text{epi}(f)$ es abierto, y por lo tanto $\text{epi}(f)$ es cerrado.
- $(v) \implies (i)$ Como $\Gamma_\gamma(f) \times \{\gamma\}$ se puede escribir como la intersección de $\text{epi}(f)$ con $\mathbf{X} \times \{\gamma\}$, deducimos que $\Gamma_\gamma(f) \times \{\gamma\}$ es cerrado en $\mathbf{X} \times \mathbb{R}$, y de aquí que $\Gamma_\gamma(f)$ es cerrado.

□

[1] La demostración es válida para espacios topológicos generales con la misma Definición 2.1.

Ejemplo 2.2.1. *Consideremos la función* $f:\mathbb{R}\to\mathbb{R}\cup\{+\infty\}$ *definida por*

$$f(x)=\begin{cases}0 & si\ x\in[-1,1]\\ +\infty & si\ no\end{cases}$$

Notemos que $epi(f)=[-1,1]\times[0,+\infty)$*, este último siendo un conjunto cerrado de* $\mathbb{R}^2$*, implica que* f *es s.c.i.. Notemos además que* $\Gamma_\gamma(f)=[-1,1]$ *si* $\gamma\geq 0$ *y* $\Gamma_\gamma(f)=\emptyset$ *si* $\gamma<0$*, siendo en ambos casos conjuntos cerrados de* $\mathbb{R}$*.*

2.3. Compacidad

El segundo ingrediente para obtener existencia es la compacidad, que recordamos ahora.

Recuerdo : Compacidad en espacios de Hilbert

Sea $(\mathbf{X},\langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $\mathcal{T}\in\{\sigma(\mathbf{X},\mathbf{X}),\mathcal{T}_{\|\cdot\|}\}$. Un conjunto $K\subseteq\mathbf{X}$ es *compacto* en $\mathcal{T}$ si todo recubrimiento de abiertos en $\mathcal{T}$ tiene un subrecubrimiento finito o, equivalentemente, si satisface

(PIF)
$$\begin{cases}(\forall(F_\alpha)_{\alpha\in\Lambda}\text{ subconjuntos cerrados de }K\text{ para }\mathcal{T})\\ [(\forall I\subset\Lambda\text{ finito})\ \cap_{\alpha\in I}F_\alpha\neq\emptyset\ \Rightarrow\ \cap_{\alpha\in\Lambda}F_\alpha\neq\emptyset].\end{cases}$$

En general, todo conjunto compacto en $\mathcal{T}$ es cerrado en $\mathcal{T}$. Además, si K es compacto en $\mathcal{T}$ y $C\subset K$ es cerrado en $\mathcal{T}$, entonces C es compacto en $\mathcal{T}$. Por otra parte, se dice que $K\subseteq\mathbf{X}$ es *secuencialmente compacto* en $\mathcal{T}$ si y sólo si para toda sucesión $\{x_n\}_{n\in\mathbb{N}}\subseteq K$, existe una subsucesión $\{x_{n_k}\}_{k\in\mathbb{N}}$ y $x\in K$ tal que $x_{n_k}\to x$ en $\mathcal{T}$.
En particular, si $\mathcal{T}=\mathcal{T}_{\|\cdot\|}$, $K\subseteq\mathbf{X}$ es secuencialmente compacto si y sólo si K es compacto, equivalencia válida en topologías metrizables gracias al teorema de Bolzano-Weierstrass (Teorema 2.3.4 en Apunte Análisis I, MAT225). En este contexto, todo conjunto compacto es acotado y cerrado fuerte, pero la recíproca no es cierta.
Por otra parte, si $\mathcal{T}=\sigma(\mathbf{X},\mathbf{X})$, $K\subseteq\mathbf{X}$ es secuencialmente compacto si y sólo si K es compacto gracias al teorema de Eberlein-Šmulian, válido en espacios de Banach (ver Dunford-Schwartz, Parte I, p. 430). Además, ambas nociones de compacidad son equivalentes a que K sea acotado y cerrado débil, gracias al teorema de Alaoglu.

Lema 2.2 (Teorema de Alaoglu en Hilbert)**.** *Sea* $(\mathbf{X},\langle\cdot,\cdot\rangle)$ *un espacio de Hilbert. Entonces toda sucesión acotada tiene una subsucesión que converge débilmente.*

Demostración. Ver Apéndice. □

2.4. Teorema de Weierstrass-Hilbert-Tonelli I

Con estas definiciones y recuerdos en mano podemos enunciar el teorema básico de existencia de mínimos.

Teorema 2.1 (Weierstrass-Hilbert-Tonelli I)**.** *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert,* $\mathcal{T} \in \{\sigma(\mathbf{X},\mathbf{X}), \mathcal{T}_{\|\cdot\|}\}$ *y* $f : \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$ *una función propia* $\mathcal{T}$*-s.c.i. Supongamos que* $\exists \gamma_0 > \text{ínf}_{\mathbf{X}}(f)$ *tal que* $\Gamma_{\gamma_0}(f)$ *es compacto en* $\mathcal{T}$*. Entonces,* $\text{ínf}_{\mathbf{X}}(f) \in \mathbb{R}$ *y* $\arg\text{mín}_{\mathbf{X}}(f) \neq \emptyset$*.*

Demostración. Haremos dos demostraciones: una topológica y otra usando el *método directo*. Primero, sea $\bar{v} = \text{ínf}_{\mathbf{X}}(f)$ y notemos que $\bar{v} \in \mathbb{R}\cup\{-\infty\}$, puesto que f es propia.

Dem. topológica: Dado que $\Gamma_\alpha(f) \subseteq \Gamma_\beta(f)$, si $\alpha \leq \beta$, tenemos que

$$\arg\text{mín}_{\mathbf{X}}(f) = \bigcap_{\gamma\in]\bar{v},+\infty[} \Gamma_\gamma(f) = \bigcap_{\gamma\in]\bar{v},\gamma_0[} \Gamma_\gamma(f).$$

Como f es $\mathcal{T}$-s.c.i. y, para todo $\gamma \in]\bar{\nu},\gamma_0[$, $\Gamma_\gamma(f) \subset \Gamma_{\gamma_0}(f)$ se tiene que $(\Gamma_\gamma(f))_{\gamma\in]\bar{v},\gamma_0[}$ son conjuntos compactos que además, por la definición de $\bar{v}$ como ínfimo, son no vacíos. Más aún, dados $\gamma_1,\ldots,\gamma_n \in \mathbb{R}$, con $\gamma = \text{mín}\{\gamma_1,\ldots,\gamma_n\} > \bar{v}$ tenemos que

$$\bigcap_{i=1}^{n} \Gamma_{\gamma_i}(f) = \Gamma_\gamma(f) \neq \emptyset.$$

Por (PIF) concluimos que $\arg\text{mín}_{\mathbf{X}}(f) = \bigcap_{\gamma\in(\bar{v},\gamma_0)} \Gamma_\gamma(f) \neq \emptyset$. A posteriori, se concluye que si $\bar{x} \in \arg\text{mín}_{\mathbf{X}}(f)$, entonces $f(\bar{x}) \leq \text{ínf}_{x\in\mathbf{X}} f(x) \leq f(\bar{x})$, de donde $f(\bar{x}) = \bar{\nu} \in \mathbb{R}$.

Dem. con método directo: Construimos primero una sucesión $\{x_n\}_{n\in\mathbb{N}}$ minimizante para f, es decir, una sucesión tal que $f(x_n) \to \bar{\nu}$ con $x_n \in \Gamma_{\gamma_0}(f)$ para todo $n \in \mathbb{N}$. Si $\bar{\nu} > -\infty$, la definición de ínfimo implica la existencia de una sucesión $\{x_n\}_{n\in\mathbb{N}}$ de $\mathbf{X}$ tal que

$$(\forall n \in \mathbb{N}) \quad \bar{\nu} \leq f(x_n) \leq \bar{\nu} + \frac{\gamma_0 - \bar{\nu}}{n+1}.$$

Por otra parte, si $\bar{\nu} = -\infty$ tomamos $x_n \in \mathbf{X}$ tal que $f(x_n) \leq \text{mín}\{-n,\gamma_0\}$ (*a posteriori* veremos que este caso no puede ocurrir). Luego, tenemos que

$$f(x_n) \to \bar{\nu} \quad \text{y} \quad f(x_n) \leq \gamma_0.$$

En particular, $x_n \in \Gamma_{\gamma_0}(f)$, que es secuencialmente compacto en $\mathcal{T}$ (ver recuerdo de compacidad). Se sigue que podemos extraer una subsucesión $\{x_{n_k}\}_{k\in\mathbb{N}}$ y $\bar{x} \in \Gamma_{\gamma_0}(f)$ tal que $x_{n_k} \to \bar{x}$ en $\mathcal{T}$. Luego, dado $A \in \mathcal{N}_{\bar{x}}$, existe $k_0 \in \mathbb{N}$ tal que $\{x_{n_k}\}_{k\geq k_0} \subseteq A$, de donde

$$\inf_{y\in A} f(y) \leq \inf_{k\geq k_0} f(x_{n_k}) \leq \sup_{k_0\in\mathbb{N}} \inf_{k\geq k_0} f(x_{n_k}) = \liminf_{k\to+\infty} f(x_{n_k}).$$

Además, notando que $f(x_{n_k}) \to \bar{\nu}$, la semicontinuidad inferior de f en $\mathcal{T}$ y la Proposición 2.1(iii) implican que

$$\bar{\nu} \leq f(\bar{x}) \leq \sup_{A\in\mathcal{N}_{\bar{x}}} \inf_{y\in A} f(y) \leq \liminf_{k\to+\infty} f(x_{n_k}) = \lim_{k\to+\infty} f(x_{n_k}) = \bar{\nu}.$$

De aquí concluimos que $f(\bar{x}) = \bar{\nu} > -\infty$. □

Una propiedad de las funciones que se relaciona con la compacidad de los conjuntos de nivel en algunos casos es la coercividad.

Definición 2.2. *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbert. Una función* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *se dice* coerciva *si y sólo si*

$$\lim_{\|x\| \to \infty} f(x) = +\infty.$$

Ejemplo 2.4.1. *Consideremos* $p \in \mathbb{N}$ *y la función* $f_p : \mathbb{R} \to \mathbb{R}$ *dada por*

$$f_p(x) = x^p, \quad \forall x \in \mathbb{R}$$

Luego, tenemos que f_p *es coerciva si y sólo si* p *es par y* $p \neq 0$*. En efecto, si* $p = 0$*, entonces* $f_p(x) = 1$ *para todo* $x \in \mathbb{R}$ *y por lo tanto no puede ser coerciva. Por otro lado, si* $p > 0$ *es par, entonces*

$$f_p(x) = x^p = |x|^p \geq |x|, \quad \forall x \in \mathbb{R} \text{ tal que } |x| \geq 1.$$

Finalmente, si $p > 0$ *es impar tenemos que* $p - 1$ *es par*

$$f_p(x) = x^p = |x|^{p-1} x, \quad \forall x \in \mathbb{R}.$$

De aquí obtenemos que f_p *no puede ser coerciva pues*

$$\lim_{x \to -\infty} f_p(x) = -\infty.$$

De la Definición 2.2 es facil probar que todos los conjuntos de nivel de funciones coercivas son acotados. Por lo tanto, si se demuestra que la funcion es $\sigma(\mathbf{X}, \mathbf{X})$-s.c.i., todos los conjuntos de nivel serán compactos en $\sigma(\mathbf{X}, \mathbf{X})$ al ser cerrados y acotados (ver recuerdo de compacidad) y la existencia de soluciones estará garantizada por el Teorema 2.1.

Caso dimension finita

Cabe recordar que, en el caso en que $\mathbf{X}$ es de dimensión finita, la topología débil y la topología de la norma (o fuerte) son equivalentes. En efecto, la inclusión $\sigma(\mathbf{X}, \mathbf{X}^*) \subset \mathcal{T}_{\|\cdot\|}$ es clara de la definición pues con $\mathcal{T}_{\|\cdot\|}$ la familia de funcionales lineales continuos es continua y $\sigma(\mathbf{X}, \mathbf{X}^*)$ es la topología con menos abiertos que logra lo mismo. Para la inclusión recíproca, supongamos por simplicidad y sin perder generalidad que $\mathbf{X} = \mathbb{R}^n$ y $\|\cdot\| = \|\cdot\|_\infty : x \mapsto \max_{i=1,\ldots,n} |x_i|$ (pues las normas son equivalentes), sea $A \in \mathcal{T}_{\|\cdot\|}$ y sea $x \in A$. Como las bolas son una base de $\mathcal{T}_{\|\cdot\|}$, existe $\varepsilon > 0$ tal que $B_{\mathbb{R}^n}(x, \varepsilon) \subset A$. Fijando $x_1^*, \ldots, x_n^*$ como los vectores canónicos de $\mathbb{R}^n$, se tiene que si $y \in \mathbb{R}^n$ cumple

$$(\forall i \in \{1, \ldots, n\}) \quad |y_i - x_i| = |\langle x_i^*, y - x \rangle| < \varepsilon, \tag{2.1}$$

se tiene $y \in B_{\mathbb{R}^n}(x, \varepsilon) \subset A$ y luego $A \in \sigma(\mathbf{X}, \mathbf{X}^*)$. Notemos que la última inclusión es sólo válida en dimensión finita, ya que si la dimensión de $\mathbf{X}$ fuese infinita, no se puede incluir en una bola una intersección finita de "franjas" del tipo

$$\cap_{i=1}^n \{y \in \mathbf{X} \mid |\langle x_i^*, y - x \rangle| < \varepsilon\},$$

ya que es un conjunto no acotado.

2.5. Ejercicios

1. PROBLEMA DE MODELAMIENTO MATEMÁTICO

   Una fábrica realiza 3 componentes A, B y C usando la misma manera de producir para cada uno de ellos. Una unidad de A toma 1 hora en producirse, una unidad de B toma 0.75 horas en producirse y una unidad de C toma 0.5 horas. Además C debe ser terminado a mano tomando 0.25 horas por unidad. Cada semana la producción no a mano no debe sobrepasar las 300 horas y la hecha a mano no debe superar las 45 horas. Las componentes son finalmente juntadas para hacer 2 productos finales. Un producto consiste de 1 unidad de A y 1 de C, y se vende a $ 30, mientras que el otro producto consiste en 2 unidades de B y una de C, y se vende a $ 45. A lo más 130 unidades del primer producto y 100 del segundo se pueden vender cada semana. Plantee el problema de programación lineal en 2 variables y resuélvalo gráficamente.

2. PROBLEM MAX-CUT Y LA PROGRAMACIÓN SEMI-DEFINIDA

   Dado un grafo $G = (V,E)$ con pesos positivos en los arcos, el problema consiste en encontrar una colección de nodos $W \subseteq V$, de forma tal que la suma de los pesos de los arcos que tienen un extremo en W y el otro en $V \setminus W$ sea máxima.

   Sea $V = \{v_1, \ldots, v_n\}$ y supongamos que los pesos en los arcos del grafo están representadas por una matriz $C \in \mathbb{M}_{n\times n}(\mathbb{R})$ que satisface

   $$\begin{cases} C_{ij} > 0 & \text{si } (v_i, v_j) \in E \\ C_{ij} = 0 & \text{si no} \end{cases}$$

   Dado que la condición $(v_i, v_j) \in E$ es equivalente a $(v_j, v_i) \in E$, tenemos que C es una matriz simétrica. Supongamos ahora que tenemos una colección de nodos $W \subseteq V$, luego la suma de los pesos de los arcos que tienen un extremo en W y el otro en $V \setminus W$

   Consideremos ahora la variable de decisión que representa a la colección de nodos $W \subseteq V$

   $$x_i = \begin{cases} 1 & \text{si } v_i \in W, \\ -1 & \text{si } v_i \in V \setminus W, \end{cases} \qquad \forall i = 1, \ldots, n.$$

   Notemos que $x_i x_j = -1$ si y sólo si

   $$(v_i \in W \ \wedge \ v_j \in V \setminus W) \qquad \vee \qquad (v_i \in V \setminus W \ \wedge \ v_j \in W).$$

   El problema se formula como sigue

   $$\text{(P)} \qquad \text{Maximizar } \sum_{i,j=1}^{n} C_{ij} \left(\frac{1 - x_i x_j}{2} \right) \text{ sobre los } x \in \mathbb{R}^n \text{ tales que } x_i^2 = 1, \ \forall i = 1, \ldots, n.$$

   Este problema es NP-duro (es decir, es muy difícil de resolver y no se sabe si se puede resolver en tiempo polinomial), por esta razón muchas veces se prefiere resolver un problema relajado. Para esto se considera que las variables $x_1, \ldots, x_n$ ahora son vectores (no números reales)

(P_n) Maximizar $\displaystyle\sum_{i,j=1}^{n} C_{ij}\left(\frac{1-x_i^\top x_j}{2}\right)$ sobre los $x \in \mathbb{R}^n$ tales que $\|x_i\|^2 = 1,\ \forall i = 1,\ldots,n.$

El problema (P_n) parece igual de difícil que (P), pero esto no es así. De hecho, (P_n) se puede resolver en tiempo polinomial (en general de forma eficaz). En efecto este problema se puede escribir como un problema de programación lineal en el espacio $\mathbb{S}^n_+(\mathbb{R})$, de las matrices simétricas y semi-definidas positivas de dimensión n, es decir, un problema de programación semi-definida.

a) Denotemos por $\mathbb{S}^n$ el espacio de matrices simétricas de dimensión n. Muestre que la función $\langle\cdot,\cdot\rangle : \mathbb{S}^n \times \mathbb{S}^n \to \mathbb{R}$ definida por

$$\langle A, B\rangle = \operatorname{tr}(AB), \qquad \forall A, B \in \mathbb{S}^n$$

es un producto interno sobre $\mathbb{S}^n$ y que por lo tanto $(\mathbb{S}^n, \langle\cdot,\cdot\rangle)$ es un espacio de Hilbert.

b) Considere la matriz de Gram asociada a una colección de vectores $\{x_1,\ldots,x_n\}$

$$P \in \mathbb{M}_{n\times n}(\mathbb{R}) \text{ con } P_{ij} = x_i^\top x_j.$$

Muestre que $P \in \mathbb{S}^n_+(\mathbb{R})$, con $P = X^\top X$, donde $X = [x_1 \ldots x_n] \in \mathbb{M}_{n\times n}(\mathbb{R})$.

c) Formular el problema (P_n) como un problema de programación semi-definida.

3. PROPIEDADES DE FUNCIONES S.C.I.

Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $\{f_\alpha\}_{\alpha\in\Lambda}$ una familia arbitraria no vacía de funciones $\mathcal{T}$-s.c.i. definidas sobre $\mathbf{X}$, es decir, para cada $\alpha \in \Lambda$ tenemos que $f_\alpha : \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$ es $\mathcal{T}$-s.c.i., con $\mathcal{T} \in \{\mathcal{T}_{\|\cdot\|}, \sigma(\mathbf{X},\mathbf{X})\}$.

a) Pruebe que $\sup\limits_{\alpha\in\Lambda}(f_\alpha)$ es $\mathcal{T}$-s.c.i., donde

$$\sup_{\alpha\in\Lambda}(f_\alpha)(x) := \sup\{f_\alpha(x) \mid \alpha \in \Lambda\}, \quad \forall x \in \mathbf{X}.$$

Indicación: Demuestre que $\operatorname{epi}(\sup\limits_{\alpha\in\Lambda} f_\alpha) = \bigcap\limits_{\alpha\in\Lambda} \operatorname{epi}(f_\alpha)$.

b) Suponga que $\Lambda = \{\alpha_1,\ldots,\alpha_n\}$ con $n \in \mathbb{N}$ dado. Demuestre que $\min\limits_{i=1,\ldots,n} f_{\alpha_i}$ y $\sum_{i=1}^n f_{\alpha_i}$ son ambas $\mathcal{T}$-s.c.i., donde

$$\min_{i=1,\ldots,n}(f_{\alpha_i})(x) := \min\{f_{\alpha_1}(x),\ldots,f_{\alpha_n}(x)\}, \quad \forall x \in \mathbf{X}.$$

PARTE I

TEORÍA GLOBAL DE OPTIMIZACIÓN

Caso Convexo

Resumen. En esta parte del curso nos enfocaremos problemas de optimización convexa, es decir, donde todos los elementos que determinan el problema de interés (función objetivo y restricciones) satisfacen una propiedad estructural llamada *convexidad*. La optimización convexa tiene el mismo status dentro de la teoría general de optimización que las ecuaciones diferenciales lineales tienen en la teoría general de ecuaciones diferenciales, pues es la base para muchas aplicaciones ya que incluye en particular la programación lineal y los problemas cuadráticos.

CAPÍTULO 3

Convexidad

Abstract. En este capítulo introduciremos formalmente la definición de una función convexa y conjunto convexo. Presentaremos problemas clásicos y actuales de optimización convexa tanto en dimensión finita como infinita.

3.1. Introducción

Comenzamos esta parte del curso recordando la definición de un conjunto convexo en un espacio de Hilbert $(\mathbf{X}, \langle\cdot,\cdot\rangle)$[1]. Un conjunto $\mathbf{S} \subseteq \mathbf{X}$ se dice *convexo* si y sólo si

$$\lambda x + (1-\lambda)y \in \mathbf{S}, \quad \forall x, y \in \mathbf{S},\ \forall \lambda \in [0,1].$$

Además se tiene el siguiente lema de accesibilidad que se demuestra en MAT410.

Proposición 3.1. *Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $\mathbf{S} \subset \mathbf{X}$ un conjunto convexo. Entonces, para todo $x \in int\,\mathbf{S}$ e $y \in \overline{\mathbf{S}}$, se tiene*

$$(\forall \lambda \in]0,1]) \quad \lambda x + (1-\lambda)y \in int\,\mathbf{S}.$$

En particular, $int\,\mathbf{S}$ y $\overline{\mathbf{S}}$ son conjuntos convexos.

Una función $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ se dirá *convexa* si y sólo si

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \quad \forall x, y \in \mathbf{X},\ \forall \lambda \in [0,1].$$

De esta desigualdad es directo que, si $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ es convexa entonces $\mathrm{dom}(f)$ es un conjunto convexo de $\mathbf{X}$. La clase de funciones convexas es cerrada para la suma y la multiplicación por escalares positivos, y el supremo de funciones convexas es convexo (ver Ejercicio 1). Otras operaciones que preservan la convexidad pueden verse en los ejercicios del capítulo (ver, por ejemplo, Ejercicio 2).

A continuación listamos otras propiedades esenciales de funciones convexas.

Proposición 3.2. *Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función dada. Luego f es convexa si y sólo si $epi(f)$ es un conjunto convexo de $\mathbf{X} \times \mathbb{R}$. Además, si f es convexa, entonces se tiene que $dom(f)$ y $\Gamma_\gamma(f)$ son conjuntos convexos para cualquier $\gamma \in \mathbb{R}$.*

Demostración. Sean (x,μ) y (y,η) en $\mathrm{epi}(f)$ y sea $\lambda \in [0,1]$. Como f es convexa, se tiene $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \leq \lambda\mu + (1-\lambda)\eta$, donde la última desigualdad proviene de la definición de $\mathrm{epi}(f)$. Para la recíproca, basta notar que $(x, f(x))$ y $(y, f(y))$ están en $\mathrm{epi}(f)$, por lo que de la definición de conjunto convexo en $\mathbf{X} \times \mathbb{R}$ se deduce la desigualdad de convexidad. La convexidad de $\Gamma_\gamma(f)$ es directa de la definición (ejercicio). □

[1] Esta parte es válida en contextos mucho más generales como espacios vectoriales topológicos localmente convexos.

Como mencionado anteriormente, en esta parte del curso nos centraremos en problemas de optimización convexa. Nuestro problema modelo de optimización

(P) Minimizar $f(x)$ sobre todos los $x \in \mathbf{X}$ que satisfacen la restricción $x \in \mathbf{S}$

se dirá convexo si $\mathbf{S}$ es un subconjunto convexo de $\mathbf{X}$ y la función $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ es convexa.

3.2. Ejemplos de problemas convexos

3.2.1. Problemas lineales

El problema de programación lineal y programación semi-definida son ejemplos de problemas convexos. En efecto, los costos, al ser funciones lineales son también funciones convexas. Además, el conjunto de restricciones son poliedros convexos en $\mathbb{R}^n$ y $\mathbb{S}^n$, respectivamente.

3.2.2. Problema lineal cuadrático - tiempo discreto

Esta clase de problemas, el primero en dimensión infinita que mencionaremos, consiste en minimizar un funcional cuyo argumento es una sucesión generada por una regla de recurrencia lineal

$$x_{k+1} = Ax_k + Bu_k, \quad \forall k \in \mathbb{N}, \tag{3.1}$$

Donde $A \in \mathbb{M}_{n\times n}(\mathbb{R})$ y $B \in \mathbb{M}_{n\times m}(\mathbb{R})$. El problema consiste en minimizar, para ciertas matrices simétricas y definidas positivas $P \in \mathbb{S}^n_{++}(\mathbb{R})$ y $Q \in \mathbb{S}^m_{++}(\mathbb{R})$, un funcional del tipo

$$f(\{x_k\}, \{u_k\}) = \frac{1}{2} \sum_{k=0}^{\infty} \left(x_k^\top P x_k + u_k^\top Q u_k \right)$$

En este caso, el espacio natural para estudiar el problema es $\mathbf{X} = \ell_2(\mathbb{R}^n) \times \ell_2(\mathbb{R}^m)$, donde $\ell_2(\mathbb{R}^N)$ es el espacio de la sucesiones $\{x_k\}$ en $\mathbb{R}^N$ tales que

$$\sum_{k=1}^{\infty} |x_k|^2 < +\infty.$$

3.2.3. Problema lineal cuadrático - tiempo continuo

La versión en tiempo continuo del problema lineal cuadrático definido sobre un intervalo $[0,T]$ corresponde a minimizar un funcional integral

$$(x,u) \mapsto \frac{1}{2} \int_0^T \left(x(t)^\top P x(t) + u(t)^\top Q u(t) \right) dt$$

el cual queda bien definido sobre el espacio $L^2_n[0,T] \times L^2_m[0,T]$. En este caso la recurrencia lineal se transforma en una ecuación diferencial parametrizada, es decir,

$$\dot{x}(t) = Ax(t) + Bu(t), \quad \text{c.t.p. } t \in [0,T].$$

3.3. Minimización convexa

Recordemos que el teorema de Weierstrass-Hilbert-Tonelli (Teorema 2.1) requiere compacidad y semicontinuidad inferior para una misma topología. Mientras que la semicontinuidad inferior es más fácil de verificar en la topología fuerte $\mathcal{T}_{\|\cdot\|}$ que en la topología débil $\sigma(\mathbf{X},\mathbf{X})$, la compacidad es más difícil de obtener en la topología fuerte que en la débil, dado que en esta última hay menos abiertos. De ese modo, si elegimos la topología débil para facilitar la compacidad de algún conjunto de nivel, verificar directamente la semicontinuidad inferior de una función con respecto a esta topología puede ser muy difícil. Es aquí donde la convexidad juega un rol importante.

Antes de continuar con el estudio de funciones convexas y aplicaciones a la optimización, revisaremos una herramienta fundamental del Análisis Convexo, la cual se refiere a la separación de convexos: el teorema de Hahn-Banach geométrico.

Recuerdo : Teorema Geométrico de Hahn-Banach

La idea básica de la versión geométrica del teorema de Hahn-Banach es que conjuntos convexos, no vacíos y disjuntos, se pueden separar por un hiperplano. Si alguno de los conjuntos resulta ser compacto y el otro cerrado, entonces la separación puede entenderse en un sentido estricto. En la Figura 3.1 hemos bosquejado interpretaciones geométricas de este teorema. El dibujo de la izquierda muestra la separación cuando uno de los conjuntos es abierto y el dibujo de la derecha un caso de separación estricta.

Lema 3.1 (Hahn-Banach I)**.** *Sea $(\mathbf{X},\langle\cdot,\cdot\rangle)$ un espacio de Hilbert. Sean $A,B\subseteq\mathbf{X}$ dos subconjuntos convexos no vacíos y disjuntos.*

(i) Si A es abierto entonces existen $x^*\in\mathbf{X}\setminus\{0\}$ y $\alpha\in\mathbb{R}$ tal que

$$\langle x^*,a\rangle<\alpha,\ \forall a\in A \quad y \quad \langle x^*,b\rangle\geq\alpha,\ \forall b\in B.$$

(ii) Si A es cerrado y B es compacto, entonces existen $x^*\in\mathbf{X}\setminus\{0\}$, $\alpha\in\mathbb{R}$ y $\varepsilon>0$ tales que

$$\langle x^*,a\rangle\leq\alpha-\varepsilon,\ \forall a\in A \quad y \quad \langle x^*,b\rangle\geq\alpha+\varepsilon,\ \forall b\in B.$$

3.3.1. Funciones convexas, semi-continuidad inferior y existencia

En general, sabemos que un conjunto cerrado para $\sigma(\mathbf{X},\mathbf{X})$ es también un conjunto cerrado para la topología fuerte $\mathcal{T}_{\|\cdot\|}$. En consecuencia, si una función es $\sigma(\mathbf{X},\mathbf{X})$-s.c.i. entonces será s.c.i para la topología fuerte. Una consecuencia importante del teorema de Hanh-Banach (Lema 3.1) para nuestros propósitos es que, para funciones convexas, la semi-continuidad inferior para la topología fuerte es indistinguible de la semi-continuidad inferior para la topología débil $\sigma(\mathbf{X},\mathbf{X})$.

Proposición 3.3. *Sea $(\mathbf{X},\langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $f:\mathbf{X}\to\mathbb{R}\cup\{+\infty\}$ una función propia convexa y $\mathcal{T}_{\|\cdot\|}$-s.c.i.. Luego se tiene que*

$$(\forall x\in\mathbf{X})\quad f(x)=\sup\{h(x)\mid\ h:\mathbf{X}\to\mathbb{R}\text{ es una función afín continua en }\mathbf{X}\text{ tal que }h(x)\leq f(x)\}.$$

Además, f es $\sigma(\mathbf{X},\mathbf{X})$-s.c.i.

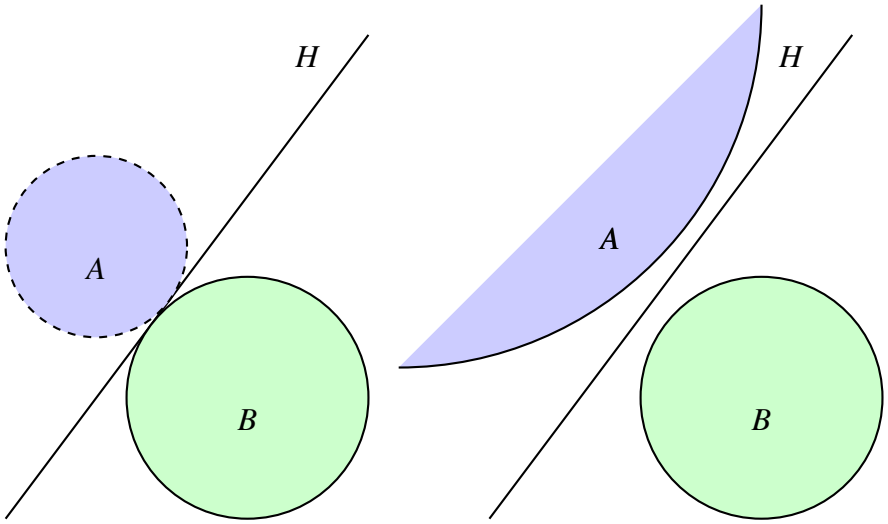


Figura 3.1: Teorema de Hahn-Banach

Demostración. Como f es convexa, s.c.i. y propia, se sigue que su epígrafo es convexo, cerrado y no vacío. Definamos

$$(\forall x \in \mathbf{X}) \quad g(x) := \sup\{h(x) \mid \ h : \mathbf{X} \to \mathbb{R} \text{ es una función afín continua en } \mathbf{X} \text{ tal que } h \leq f\}. \tag{3.2}$$

Por definición se tiene que $g \leq f$. Para demostrar la igualdad, por la definición del supremo basta probar

$$(\forall x \in \mathbf{X})(\forall r < f(x))(\exists h\colon \mathbf{X} \to \mathbb{R} \text{ afin continua en } \mathbf{X} \text{ tal que } h \leq f) \quad r < h(x). \tag{3.3}$$

Fijemos $x \in \mathbf{X}$ y separemos la demostración de (3.3) en dos partes.

1. Supongamos que $x \in \operatorname{dom}(f)$. Sea $r < f(x) < +\infty$. Luego $(x,r) \notin \operatorname{epi}(f)$ y gracias al Teorema Geométrico de Hahn-Banach (Lema 3.1), existen $(x^*,s) \in \mathbf{X} \times \mathbb{R} \setminus \{(0,0)\}$ y $\alpha \in \mathbb{R}$ tales que

   $$\langle x^*, x\rangle + sr < \alpha \leq \langle x^*, y\rangle + s\lambda, \quad \forall (y,\lambda) \in \operatorname{epi}(f). \tag{3.4}$$

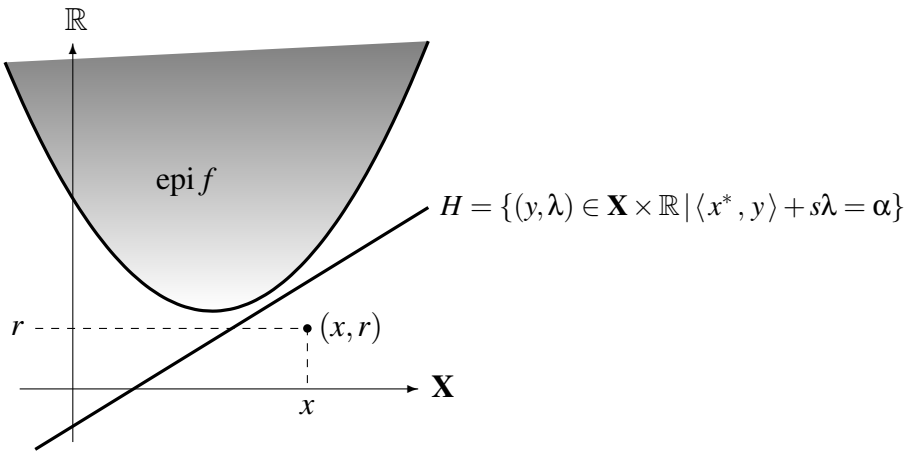
   En particular, como $(x, f(x)) \in \operatorname{epi}(f)$, concluimos de (3.4) que $s(r - f(x)) < 0$ de donde $s > 0$.

   Dividiendo en (3.4) por $s > 0$, se obtiene

   $$\frac{1}{s}\langle x^*, x - y\rangle + r < \underbrace{\frac{\alpha}{s} - \frac{1}{s}\langle x^*, y\rangle}_{h(y)} \leq \lambda, \quad \forall (y,\lambda) \in \operatorname{epi}(f), \tag{3.5}$$

   y definiendo la función afín $h\colon y \mapsto \frac{\alpha}{s} - \frac{1}{s}\langle x^*, y\rangle$, se concluye que $h \leq f$ y que $r < h(x)$.

2. Supongamos que $x \notin \operatorname{dom}(f)$y sea $r < +\infty = f(x)$. Se tiene que $(x,r) \notin \operatorname{epi}(f)$ y, al igual que en la parte anterior, por Hahn-Banach se tiene (3.4). Notemos que si $s < 0$ tomando $(y,\lambda) \in \operatorname{epi}(f)$



y haciendo $\lambda \to +\infty$ se contradice (3.4). Si $s > 0$, la función afín $h\colon y \mapsto \frac{\alpha}{s} - \frac{1}{s}\langle x^*, y\rangle$ satisface (3.3) al igual que en la parte anterior. Finalmente, si $s = 0$, tenemos que

$$\langle x^*, x\rangle < \alpha \leq \langle x^*, y\rangle, \quad \forall y \in \mathrm{dom}(f).$$

Sea $\tilde{h} : \mathbf{X} \to \mathbb{R}$ una función afín continua tal que $\tilde{h} \leq f$, cuya existencia está garantizada por la primera parte pues f es propia. Luego para todo $k \in \mathbb{N}$ e $y \in \mathbf{X}$ se tiene que

$$f(y) \geq \tilde{h}(y) \geq h_k(y) := \tilde{h}(y) + k(\alpha - \langle x^*, y\rangle),$$

pero $h_k(x) \to +\infty$ cuando $k \to +\infty$. Por lo tanto $f(x) = g(x) = +\infty$.

Finalmente, notemos que (ver Ejercicio 3 - Capítulo 1)

$$\mathrm{epi}(g) = \bigcap_{\{h:\mathbf{X}\to\mathbb{R} \text{ afín continua con } h\leq f\}} \mathrm{epi}(h).$$

Ahora bien, dado que el epígrafo de una función afín continua es cerrado para la topología débil $\sigma(\mathbf{X},\mathbf{X})$, se tiene entonces que $\mathrm{epi}(g)$ es cerrado para la topología débil $\sigma(\mathbf{X},\mathbf{X})$. En otras palabras, g es $\sigma(\mathbf{X},\mathbf{X})$-s.c.i. lo cual termina la demostración. □

En vista del resultado anterior podemos presentar una nueva versión del teorema de existencia de mínimos de Weierstrass-Hilbert-Tonelli, especializada para el caso convexo.

Teorema 3.1. *[Weierstrass-Hilbert-Tonelli II] Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función propia convexa y* $\mathcal{T}_{\|\cdot\|}$*-s.c.i. Supongamos que* $\exists \gamma_0 > \inf_{\mathbf{X}}(f)$ *tal que* $\Gamma_{\gamma_0}(f)$ *es acotado. Entonces,* $\inf_{\mathbf{X}}(f) \in \mathbb{R}$ *y* $\arg\min_{\mathbf{X}}(f) \neq \emptyset$.

Demostración. Como f es convexa y $\mathcal{T}_{\|\cdot\|}$-s.c.i., la Proposición 3.3 implica que f es $\sigma(\mathbf{X},\mathbf{X})$-s.c.i., de donde $\Gamma_{\gamma_0}(f)$ es cerrado débil. Como además $\Gamma_{\gamma_0}(f)$ es acotado por hipótesis, es compacto débil (ver recuerdo de compacidad). Por lo tanto, aplicando el Teorema 2.1 se concluye el resultado. □

En el teorema anterior la convexidad juega un rol esencial, pues permite conectar la semi-continuidad inferior para las topologías fuerte y débil.

3.3.2. Unicidad de minimizadores

Hasta el momento hemos hablado de existencia de minimizadores, pero no hemos mencionado cuántos pueden haber. Veremos ahora que en optimización convexa hay solo tres posibilidades: (i) hay una cantidad infinita no numerable de minimizadores, (ii) existe un única solución óptima, o bien (iii) no hay solución del todo. Esto es consecuencia directa de la siguiente proposición.

Proposición 3.4. *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función propia y convexa. El conjunto de minimizadores de* f*:*

$$\arg\min_{\mathbf{X}}(f) := \{\bar{x} \in \mathbf{X} \mid \ f(\bar{x}) \leq f(x), \quad \forall x \in \mathbf{X}\}$$

es convexo. Más aún, si suponemos además que f *es* estrictamente convexa*, es decir,*

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \quad \forall x, y \in \mathbf{X},\ x \neq y,\ \forall \lambda \in (0,1).$$

entonces $\arg\min_{\mathbf{X}}(f)$ *contiene a lo más un único elemento.*

Demostración. Sean $\bar{x}$ e $\bar{y}$ en $\arg\min_{\mathbf{X}}(f)$ y sea $\lambda \in [0,1]$. Entonces, para todo $z \in \mathbf{X}$, por convexidad y definición de mínimo se tiene

$$f(\lambda\bar{x} + (1-\lambda)\bar{y}) \leq \lambda f(\bar{y}) + (1-\lambda) f(\bar{x}) \leq \lambda f(z) + (1-\lambda) f(z) = f(z),$$

de donde $\lambda\bar{x} + (1-\lambda)\bar{y} \in \arg\min_{\mathbf{X}}(f)$, y luego $\arg\min_{\mathbf{X}}(f)$ es convexo. Para la unicidad, si asumimos que $\bar{x} \neq \bar{y}$, como se tiene $f(\bar{x}) = f(\bar{y})$, la convexidad estricta implica

$$f(\lambda\bar{x} + (1-\lambda)\bar{y}) < \lambda f(\bar{y}) + (1-\lambda) f(\bar{x}) = f(\bar{x}) = f(\bar{y}),$$

por lo que ni $\bar{x}$ ni $\bar{y}$ pueden ser mínimos, lo que nos lleva a una contradicción y a la conclusión. □

Notemos que el teorema anterior implica que en el caso de haber más de un mínimo, digamos $\bar{x}_1$ y $\bar{x}_2$, entonces todos los elementos del segmento

$$[\bar{x}_1, \bar{x}_2] := \{\lambda\bar{x}_1 + (1-\lambda)\bar{x}_2 \mid \ \lambda \in [0,1]\}$$

son también mínimos, lo que implica que $\arg\min_{\mathbf{X}}(f)$ es un conjunto infinito no numerable.

3.4. Ejercicios

1. ÁLGEBRA DE FUNCIONES CONVEXAS Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $\{f_\alpha\}_{\alpha\in\Lambda}$ una familia arbitraria no vacía de funciones convexas definidas sobre $\mathbf{X}$, es decir, para cada $\alpha \in \Lambda$, $f_\alpha : \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$ es convexa.

   a) Pruebe que $\sup_{\alpha\in\Lambda}(f_\alpha)$ es convexa.

   b) Suponga que $\Lambda = \{\alpha_1,\ldots,\alpha_n\}$ con $n \in \mathbb{N}$ dado. Demuestre que para todo $\mu_1,\ldots,\mu_n \geq 0$ se tiene que $\sum_{i=1}^n \mu_i f_{\alpha_i}$ es una función convexa.

2. Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ e $(\mathbf{Y}, \langle\cdot,\cdot\rangle)$ espacios de Hilbert, sea $\varphi\colon \mathbf{X}\times\mathbf{Y} \to \mathbb{R}\cup\{+\infty\}$ una función convexa y definamos

$$f\colon \mathbf{X} \to \mathbb{R}\cup\{+\infty\}\colon x \mapsto \inf_{y\in\mathbf{Y}} \varphi(x,y).$$

   Demuestre que f es convexa.

3. CRITERIOS ALTERNATIVOS DE CONVEXIDAD

   Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$ una función propia.

   a) Demuestre que f es convexa si y sólo si para todo $x_1,\ldots,x_n \in \mathbf{X}$ y $\lambda_1,\ldots,\lambda_n \in [0,1]$ se tiene que

$$\sum_{i=1}^n \lambda_i = 1 \quad\Longrightarrow\quad f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f(x_i)$$

   b) Suponga que $\mathbf{X} = \mathbb{R}$ y sea $f : \mathbb{R} \longrightarrow \mathbb{R}$ una función continua que satisface la desigualdad siguiente:

$$f(x) \leq \frac{1}{2h}\int_{x-h}^{x+h} f(y)dy, \quad x \in \mathbb{R}, \quad h > 0.$$

   Pruebe:

   1) El máximo de f en un intervalo cerrado $[a,b]$ es alcanzado en a o en b.
   2) f es convexa.

      Indicación: Considere $L(x) = \dfrac{(x-a)f(b)-(x-b)f(a)}{b-a}$ y muestre que $f(x) \leq L(x)$.

4. FUNCIÓN CUADRÁTICA

   Sean $A \in \mathbb{S}^n(\mathbb{R})$, $b \in \mathbb{R}^n$ y $c \in \mathbb{R}$. Considere la función cuadrática $f : \mathbb{R}^n \to \mathbb{R}$ definida por

$$f(x) = \frac{1}{2}x^\top Ax + b^\top x + c, \quad \forall x \in \mathbb{R}^n.$$

   - Muestre que si f es acotada inferiormente, entonces $A \in \mathbb{S}^n_+(\mathbb{R})$. Muestre además que f es convexa (usando el criterio algebraico) y que además alcanza su mínimo en $\mathbb{R}^n$.
   - Pruebe que f es estrictamente convexa si y sólo si $A \in \mathbb{S}^n_{++}(\mathbb{R})$.

5. FUNCIONES MARGINALES

   Sean $\mathbf{X}$ e $\mathbf{Y}$ dos espacios vectoriales. Considere $A \subseteq \mathbf{X}$ y $B \subseteq \mathbf{Y}$ dos conjuntos convexos no vacíos. Sea $\varphi : \mathbf{X} \times \mathbf{Y} \to \mathbb{R} \cup \{+\infty\}$, una función convexa tal que $\text{ínf}\{\varphi(x,y) \mid \ y \in B\} > -\infty$ para todo $x \in A$. Pruebe que la función $f(x) = \text{ínf}\{\varphi(x,y) \mid \ y \in B\} + \delta_A(x)$ es convexa en $\mathbf{X}$.

6. PROYECCIÓN SOBRE UN CERRADO

   Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y sea $\mathbf{S} \subseteq \mathbf{X}$ un subconjunto dado. Definimos la distancia de un punto $x \in \mathbf{X}$ a $\mathbf{S}$ via la fórmula:

$$\text{dist}(x, \mathbf{S}) = \text{ínf}\{\|x - s\| \mid \ s \in \mathbf{S}\}.$$

   Definimos también el conjunto proyección sobre $\mathbf{S}$ como sigue

$$\text{proy}(x, \mathbf{S}) = \{s \in \mathbf{S} \mid \ \text{dist}(x, \mathbf{S}) = \|x - s\|\}.$$

   a) Muestre que $x \mapsto \text{dist}(x, \mathbf{S})$ es Lipschitz continua de constante $L = 1$.

   b) Pruebe que si $\mathbf{S}$ es cerrado para $\sigma(\mathbf{X}, \mathbf{X})$, la topología débil de $\mathbf{X}$, entonces el ínfimo en la definición de $\text{dist}(x, \mathbf{S})$ se alcanza y además $\text{proy}(x, \mathbf{S})$ es no vacío para todo $x \in \mathbf{X}$.

   c) Pruebe que $\mathbf{S}$ es convexo si y sólo si $x \mapsto \text{dist}(x, \mathbf{S})$ es convexa.

   d) Muestre que si $\mathbf{S}$ es convexo y cerrado (para la topología fuerte) entonces $\text{proy}(x, \mathbf{S}) \neq \emptyset$.

   e) Demuestre que

$$\text{proy}(x, \mathbf{S}) = \left\{ s \in \mathbf{S} \mid \langle y - s, x - s \rangle \leq \frac{1}{2}\|y - s\|^2, \ \forall y \in \mathbf{S} \right\}$$

   f) Muestre que si $\mathbf{S}$ es convexo, entonces $\text{proy}(x, \mathbf{S})$ tiene un único elemento y que

$$\text{proy}(x, \mathbf{S}) = \{s \in \mathbf{S} \mid \langle y - s, x - s \rangle \leq 0, \ \forall y \in \mathbf{S}\}.$$

CAPÍTULO 4

Optimización convexa diferenciable

Abstract. En este capítulo estudiaremos funciones convexas diferenciables, condiciones de optimalidad e introduciremos algunos métodos iterativos para encontrar sus mínimos. Haremos especial énfasis en problemas cuadráticos.

La convexidad de una función es un criterio algebraico, que puede ser difícil de probar algunas veces. Comenzaremos este capítulo indicando algunos criterios alternativos para las funciones diferenciables y de paso recordemos algunas definiciones básica del cálculo diferencial.

A lo largo de este capítulo trabajaremos básicamente con funciones $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ convexas tales que $\operatorname{dom}(f)$ será un abierto de un espacio de Hilbert $(\mathbf{X}, \langle \cdot, \cdot \rangle)$.

4.1. Criterios de primer orden

Estudiaremos ahora algunos criterios de primer orden que nos ayudarán a determinar si una función es convexa o no. Haremos esto usando la noción de función Gâteaux diferenciable.

Recuerdo : Funciones Gâteaux diferenciables

Supongamos que $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ es una función tal que $\operatorname{dom}(f)$ tiene interior no vacío. Diremos que la función f es Gâteaux diferenciable en $x \in \operatorname{int}(\operatorname{dom}(f))$ si

$$\lim_{t \to 0} \frac{f(x+td) - f(x)}{t} = \ell(d), \quad \forall d \in \mathbf{X},$$

donde $\ell : \mathbf{X} \to \mathbb{R}$ es un funcional lineal continuo, que se conoce como la derivada de Gâteaux de f, la cual admite un representante de Riesz conocido con el nombre de gradiente y denotado por $\nabla f(x)$, i.e., para todo $d \in \mathbf{X}$, $\ell(d) = \langle \nabla f(x), d \rangle$. Además, si $\mathbf{X} = \mathbb{R}^n$, entonces el gradiente de f puede ser representado a través de las derivadas parciales de f, es decir,

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right).$$

Teorema 4.1. *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función Gâteaux diferenciable en $\operatorname{dom}(f)$, el cual asumimos es un conjunto convexo abierto de $\mathbf{X}$. Las siguientes afirmaciones son equivalentes:*

(i) $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *es convexa.*

(ii) f *es subdiferenciable, es decir, para todo $x, y \in \operatorname{dom}(f)$, se tiene $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$.*

(iii) ∇f *es monótono, es decir, para todo* $x, y \in dom(f)$ *se tiene* $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$.

Demostración. Dividamos la demostración en cuatro partes:

$(i) \Rightarrow (ii)$ Sean $x, y \in \mathrm{dom}(f)$ y $t \in (0,1)$. De la convexidad de f se deduce

$$\frac{f(x+t(y-x)) - f(x)}{t} \leq f(y) - f(x).$$

Luego, haciendo $t \to 0$ obtenemos (ii).

$(ii) \Rightarrow (iii)$ Sean $x, y \in \mathrm{dom}(f)$. Usando (ii) y luego intercambiando los roles de x e y en la desigualdad se tienen

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

Finalmente, sumando ambas desigualdades se obtiene el resultado.

$(iii) \Rightarrow (i)$ Dados $x, y \in \mathrm{dom}(f)$ fijos. En vista que $\mathrm{dom}(f)$ es abierto, podemos escoger $\varepsilon > 0$ tal que $x + t(y-x) \in \mathrm{dom}(f)$ para cualquier $t \in (-\varepsilon, 1+\varepsilon)$. Definamos $\phi \colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ via la fórmula

$$\phi(t) := \begin{cases} f(x+t(y-x)) & \text{si } t \in (-\varepsilon, 1+\varepsilon) \\ +\infty & \text{si no.} \end{cases}$$

Como f es Gâteaux diferenciable en $\mathrm{dom}(f)$, se tiene que ϕ también lo es en su dominio. En particular, ϕ es derivable en $(-\varepsilon, 1+\varepsilon)$ y por lo tanto continua en $[0,1]$. Además, se tiene que $\phi'(t) = \langle \nabla f(x+t(y-x)), y - x \rangle$ para cualquier $t \in (-\varepsilon, 1+\varepsilon)$. Notemos que si $-\varepsilon < s < t < 1+\varepsilon$ se tiene que

$$\phi'(t) - \phi'(s) = \langle \nabla f(z_t), y - x \rangle - \langle \nabla f(z_s), y - x \rangle = \frac{1}{t-s} \langle \nabla f(z_t) - \nabla f(z_s), z_t - z_s \rangle \geq 0,$$

donde $z_t := x + t(y-x)$ y $z_t := x + s(y-x)$, y por lo tanto ϕ' es no decreciente en el intervalo $(-\varepsilon, 1+\varepsilon)$. Luego se tiene por teorema del valor medio que

$$(\forall t \in]0,1[)(\exists t^* \in]0,t[) \quad \frac{\phi(t) - \phi(0)}{t} = \phi'(t^*) \leq \phi'(t).$$

Por lo tanto, si definimos $\varphi \colon]0,1[\to \mathbb{R} \colon t \mapsto (\phi(t) - \phi(0))/t$, se tiene que φ es diferenciable en $]0,1[$ y

$$(\forall t \in]0,1[) \quad \varphi'(t) = \frac{\phi'(t) - \frac{\phi(t) - \phi(0)}{t}}{t} \geq 0,$$

de donde φ es no decreciente. Finalmente, la convexidad se deduce de que, para todo $t \in]0,1[$,

$$\frac{f(x+t(y-x)) - f(x)}{t} = \varphi(t) \leq \varphi(1) = f(y) - f(x).$$

□

Ejemplo 4.1.1. *Usando la subdiferenciablidad podemos probar fácilmente que* $x \mapsto \exp(x)$ *es una función convexa. Notemos que la desigualdad de la subdiferenciabilidad es*

$$\exp(y) \geq \exp(x) + \exp(x)(y - x)$$

y se puede re-escribir, fijando $z = y - x$ como

$$\exp(z) \geq 1 + z.$$

Esta última siendo una desigualdad fundamental de la función exponencial estudiada en cursos básicos de cálculo.

Ejemplo 4.1.2. *Usando ahora la monotonía podemos probar fácilmente que $x \mapsto -\log(x)$ es una función convexa. Notemos primero que $dom(\log) = (0, +\infty)$ y que la desigualdad de la monotonía es*

$$\left(-\frac{1}{x} + \frac{1}{y}\right)(x - y) \geq 0$$

la que podemos re-escribir como

$$\frac{(x-y)^2}{xy} \geq 0$$

la cual siempre es válida si $x, y > 0$.

4.1.1. Comentarios sobre la diferenciabilidad en el sentido de Gâteaux

En el caso $\mathbf{X} = \mathbb{R}$, se tiene que una función es Gâteaux diferenciable si y sólo si la función es derivable, y por lo demás continua. En general, si $\mathbf{X} \neq \mathbb{R}$ la diferenciabilidad en el sentido de Gâteaux no implica continuidad de una función. Por ejemplo, la función $f : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$f(x,y) = \begin{cases} 1 & \text{si } x, y > 0 \ \wedge \ x^2 > y \\ 0 & \text{si no} \end{cases}$$

es Gâteaux diferenciable en $(0,0)$, con $\nabla f(0,0) \equiv 0$, pero f no es continua en $(0,0)$. Esto constituye una de la mayores diferencias entre la diferenciabilidad en el sentido de Gâteaux y Fréchet.

Recuerdo : Funciones Fréchet diferenciables

Una función $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ se dice Fréchet diferenciable en $x \in \text{int}(\text{dom}(f))$ si es Gâteaux diferenciable y su diferencial $\nabla f(x) \in \mathbf{X}$ satisface

$$\lim_{h \to 0} \frac{|f(x+h) - f(x) - \langle \nabla f(x), h \rangle|}{\|h\|} = 0.$$

Cuando la derivada de Gâteaux es continua se puede concluir que la función es Fréchet diferenciable, como asegura el siguiente resultado.

Proposición 4.1. *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y sea $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función Gâteaux diferenciable en una vecindad de $x \in \mathbf{X}$ tal que $\nabla f : \mathbf{X} \to \mathbf{X}$ es continuo en x. Entonces f es Fréchet diferenciable en x y su derivada de Fréchet es $\nabla f(x)$.*

Demostración. Sea $\varepsilon > 0$ tal que f es Gâteaux diferenciable en $B_{\mathbf{X}}(x, \varepsilon)$, sea $h \in B_{\mathbf{X}}(x, \varepsilon)$ y definamos $\phi : t \mapsto f(x + th)$. Por Teorema del Valor Medio en $\mathbb{R}$ se tiene que existe $t \in]0,1[$ tal que $f(x+h) -$

$f(x) = \phi(1) - \phi(0) = \phi'(t) = \langle \nabla f(x+th), h \rangle$. Luego, usando la desigualdad de Cauchy-Schwarz se obtiene

$$\frac{|f(x+h) - f(x) - \langle \nabla f(x), h \rangle|}{\|h\|} \leq \|\nabla f(x+th) - \nabla f(x)\| \to 0$$

cuando $\|h\| \to 0$ por la continuidad de ∇f, lo que concluye el resultado. □

4.2. Criterios de orden superior

Veremos a continuación un criterio de orden superior para determinar la convexidad de una función. Antes de continuar recordemos algunas nociones de derivadas de orden superior.

Recuerdo : Derivadas de orden superior

Una función $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ se dice dos veces Gâteaux diferenciable en $x \in \text{int}(\text{dom}(f))$ si f Gâteaux diferenciable en una vecindad de x y además existe un operador bilineal continuo y simétrico $B : \mathbf{X} \times \mathbf{X} \to \mathbb{R}$ tal que

$$(\forall h, k \in \mathbf{X}) \quad \lim_{t \to 0} \frac{\langle \nabla f(x+th) - \nabla f(x), k \rangle}{t} = B(h,k).$$

Este funcional bilineal continuo se conoce como el diferencial de Gâteaux de segundo orden de f en x y se denota como $\nabla^2 f(x)$. Es importante mencionar que en el caso $\mathbf{X} = \mathbb{R}^n$ se tiene que $\nabla^2 f(x)$ puede ser representado a través de la matriz Hesiana de f:

$$Hf(x) = \begin{pmatrix} \partial^2_{x_1,x_1} f(x) & \partial^2_{x_1,x_2} f(x) & \dots & \partial^2_{x_1,x_n} f(x) \\ \partial^2_{x_2,x_1} f(x) & \partial^2_{x_2,x_2} f(x) & \dots & \partial^2_{x_2,x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2_{x_n,x_1} f(x) & \partial^2_{x_n,x_2} f(x) & \dots & \partial^2_{x_n,x_n} f(x) \end{pmatrix}$$

a través de la relación

$$(\forall h, k \in \mathbb{R}^n) \quad \nabla^2 f(x)(h,k) = h^\top Hf(x) k.$$

Por otro lado, f se dice dos veces Fréchet diferenciable en x si f es dos veces Gâteaux diferenciable en x y se satisface

$$\lim_{h \to 0} \sup_{\|k\|=1} \frac{|\langle \nabla f(x+h) - \nabla f(x), k \rangle - \nabla^2 f(x)(h,k)|}{\|h\|} = 0.$$

Teorema 4.2. *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función dos veces Gâteaux diferenciable en* $\text{dom}(f)$*, este último siendo un convexo abierto de* $\mathbf{X}$*. Entonces,* $f : \mathbf{X} \to$

$\mathbb{R}\cup\{+\infty\}$ *es convexa si y sólo si el operador* $\nabla^2 f$ *es semi-definido positivo, es decir,*

$$\nabla^2 f(x)(h,h) \geq 0, \quad \forall x \in dom(f),\ \forall h \in \mathbf{X}.$$

Demostración. Supongamos primero que f es convexa. Sean $x \in \mathrm{dom}(f)$, $h \in \mathbf{X}$ y $t > 0$ tal que $x+th \in \mathrm{dom}(f)$, cuya existencia está garantizada pues $\mathrm{dom}(f)$ es abierto. Del Teorema 4.1, se tiene

$$\langle \nabla f(x+th) - \nabla f(x), h\rangle = \frac{1}{t}\langle \nabla f(x+th) - \nabla f(x), x+th-x\rangle \geq 0.$$

Luego, dividiendo por t y haciendo $t \to +\infty$ llegamos a que $D^2 f(x)$ es semi-definido positivo.

Veamos ahora la implicancia recíproca. Supongamos ahora que $D^2 f$ es semi-definido positivo y sean $x, y \in \mathrm{dom}(f)$. Usando el mismo argumento que en la demostración de $[(iii) \Rightarrow (i)]$ del Teorema 4.1, podemos escoger $\varepsilon > 0$ tal que la función $\phi\colon \mathbb{R} \to \mathbb{R}\cup\{+\infty\}$ dada por

$$\phi(t) := \begin{cases} f(x+t(y-x)) & \text{si } t \in (-\varepsilon, 1+\varepsilon) \\ +\infty & \text{si no,} \end{cases}$$

es derivable en $(-\varepsilon, 1+\varepsilon)$. De hecho, dado que f es dos veces Gâteaux diferenciable en $\mathrm{dom}(f)$ se tiene que ϕ es dos veces derivable con ϕ' continua en $[0,1]$. Usando la regla de la cadena se obtiene que $\phi''(t) = \nabla^2 f(x+t(y-x))(y-x, y-x)$. Luego, como $\nabla^2 f$ es semi-definido positivo se tiene que ϕ' es no decreciente, y por lo tanto, la conclusión sigue usando los mismos argumentos que en la demostración de $[(iii) \Rightarrow (i)]$ del Teorema 4.1. □

Una ligera modificación de la demostración del resultado anterior permite obtener una condición necesaria para que una función sea estrictamente convexa.

Teorema 4.3. *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$ *una función de clase* $\mathcal{C}^2$ *en* $dom(f)$*, es último siendo un abierto de* $\mathbf{X}$*. Si el operador* $\nabla^2 f$ *es definido positivo, es decir,*

$$\nabla^2 f(x)(h,h) > 0, \quad \forall x \in dom(f),\ \forall h \in \mathbf{X}\setminus\{0\}.$$

entonces $f : \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$ *es estrictamente convexa.*

Demostración. Ejercicio. □

Notemos que el converso del teorema 4.3 no es válido, de hecho la función $x \mapsto x^4$ es estrictamente convexa, pero su segunda derivada en $x = 0$ es nula.

Ejemplo 4.2.1. *Consideremos la función cuadrática* $f : \mathbb{R}^n \to \mathbb{R}$ *definida via*

$$f(x) = \frac{1}{2}x^\top Ax - b^\top x + c, \quad \forall x \in \mathbb{R}^n.$$

Donde $A \in \mathbb{S}^n$*,* $b \in \mathbb{R}^n$ *y* $c \in \mathbb{R}$*. Luego se tiene que* $\nabla^2 f(x) = A$ *y por lo tanto tenemos que* f *es convexa si y sólo si* $A \in \mathbb{S}^n_+(\mathbb{R})$*. Notemos que si* $A \in \mathbb{S}^n_{++}(\mathbb{R})$ *entonces* f *es estrictamente convexa. En este caso particular (y no en general) se tiene también que el converso es cierto, es decir, si* f *es estrictamente convexa entonces* $\nabla^2 f(x) = A \in \mathbb{S}^n_{++}(\mathbb{R})$ *(ver Ejercicio 4 - Capítulo 3).*

4.3. Regla de Fermat

En vista del Teorema 4.1, tenemos una forma fácil de caracterizar mínimos de una función convexa Gâteaux diferenciable, la cual se resume en el siguiente resultado.

Teorema 4.4 (Regla de Fermat I)**.** *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función convexa Gâteaux diferenciable en* $\mathrm{dom}(f)$*, este último siendo un abierto de* $\mathbf{X}$*. Luego* $\bar{x} \in \mathbf{X}$ *es un mínimo de* f *si y sólo si* $\nabla f(\bar{x}) = 0$*.*

Demostración. Si $\bar{x} \in \mathbf{X}$ es un mínimo de f, entonces

$$(\forall h \in \mathbf{X})(\forall t \neq 0) \quad \frac{f(\bar{x}+th) - f(\bar{x})}{t} \geq 0 \tag{4.1}$$

y pasando al límite cuando $t \to 0$ se concluye $\langle \nabla f(\bar{x}), h \rangle \geq 0$. Tomando $h = -\nabla f(\bar{x})$ se deduce $\nabla f(\bar{x}) = 0$. La recíproca se concluye del Teorema 4.1. □

Es importante mencionar que en el caso convexo la condición $\nabla f(\bar{x}) = 0$ es suficiente y necesaria para que $\bar{x}$ sea un mínimo. En problemas no convexo, incluso sin restricciones, esto no es, en general, cierto. Puntos que satisfacen la condición $\nabla f(\bar{x}) = 0$ son llamados *puntos críticos de* f.

4.3.1. Aplicación a problemas cuadráticos

Retomando lo visto en Ejemplo 4.2.1 tenemos que si $A \in \mathbb{S}^n_+(\mathbb{R})$ entonces la función $f : \mathbb{R}^n \to \mathbb{R}$ dada por

$$f(x) = \frac{1}{2} x^\top A x - b^\top x + c, \quad \forall x \in \mathbb{R}^n,$$

es convexa. Luego aplicando la Regla de Fermat se tiene que $\bar{x}$ es un mínimo de f si y sólo si la ecuación

$$A\bar{x} = b$$

tiene solución, es decir, si $b \in \mathrm{im}(A)$. En particular, si A no es invertible entonces f puede tener infinitas soluciones (si $b \in \mathrm{im}(A)$) o bien ninguna si $b \notin \mathrm{im}(A)$.

Ejemplo 4.3.1. *Consideremos* $c = 0$*,* $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ *y* $b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$*. Notemos que* $b \notin im(A)$*. Por otro lado*

$$f(x_1, x_2) = x_1^2 - x_2$$

Por lo tanto $f(0,k) \to -\infty$ *si* $k \to +\infty$*. De donde concluimos que* f *no admite un mínimo.*

Caso estrictamente convexo

Notemos que si $A \in \mathbb{S}^n_{++}(\mathbb{R})$ entonces A es invertible y $\bar{x} = A^{-1}b$. Esto se condice con el hecho que f es estrictamente convexa y que por lo tanto su mínimo es único. Además, la existencia de mínimo también está asegurada por el teorema de Weierstrass-Hilbert-Tonelli pues f es coerciva como veremos a continuación.

Proposición 4.2. *Sean $A \in \mathbb{S}^n_{++}(\mathbb{R})$, $b \in \mathbb{R}^n$ y $c \in \mathbb{R}$. La función*

$$f(x) = \frac{1}{2}x^\top Ax - b^\top x + c$$

es coerciva. Más aún, si $\lambda > 0$ es el menor valor propio de A entonces

$$\lambda|x|^2 \leq x^\top Ax$$

Demostración. Como la matriz A es simétrica, admite una descomposición del tipo $A = PDP^\top$ con D la matriz diagonal con los valores propios reales $\lambda_1 \geq \cdots \geq \lambda_n$ de A y P la matriz cuyas columnas son los vectores propios ortonormales $v_1, \ldots, v_n$ asociados a los valores propios $\lambda_1, \ldots, \lambda_n$; notar que $PP^\top = P^\top P = I$, con I siendo la matriz identidad. Además, como A es definida positiva, todos sus valores propios son (estrictamente) positivos. Más aún, como $v_1, \ldots, v_n$ constituyen una base de $\mathbb{R}^n$, para todo $x \in \mathbb{R}^n$, existen coeficientes reales $\xi_1, \ldots, \xi_n$ tales que

$$x = \sum_{i=1}^n \xi_i v_i = Py, \qquad \text{donde } y = (\xi_1, \ldots, \xi_n).$$

De este modo, $|x|^2 = x^\top x = (Py)^\top Py = y^\top P^\top Py = y^\top y = |y|^2$. Por lo tanto

$$(Ax)^\top x = (PDP^\top x)^\top x = (DP^\top x)^\top (P^\top x) = (Dy)^\top y = \sum_{i=1}^n \xi_i^2 \lambda_i \geq \lambda_n |y|^2 = \lambda_n |x|^2,$$

de donde se obtiene la coercividad. □

4.4. Principio Variacional de Ekeland

El Principio Variacional de Ekeland permite construir una Regla de Fermat aproximada en la ausencia de coercividad. En tal caso, el hecho que la función objetivo sea acotada inferiormente es importante como queda demostrado con el Ejemplo 4.3.1.

Teorema 4.5. *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ es un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función propia, convexa y s.c.i para la topología generada por la norma e inferiormente acotada. Consideremos $\varepsilon > 0$, $\lambda > 0$ y sea $x_0 \in \mathbf{X}$ tal que*

$$f(x_0) \leq \text{ínf}_{\mathbf{X}}(f) + \varepsilon. \tag{4.2}$$

Entonces, existe un punto $x_\varepsilon \in \mathbf{X}$ tal que

(i) $f(x_\varepsilon) \leq f(x_0)$,

(ii) $\|x_\varepsilon - x_0\| \leq \lambda$,

(iii) $f(x_\varepsilon) < f(x) + \frac{\varepsilon}{\lambda}\|x - x_\varepsilon\|$, *para todo* $x \in \mathbf{X} \setminus \{x_\varepsilon\}$.

Si además f Gâteaux diferenciable en $\text{dom}(f)$, el cual asumimos es un abierto de $\mathbf{X}$, entonces $\|\nabla f(x_\varepsilon)\| \leq \frac{\varepsilon}{\lambda}$ y existe una sucesión minimizante $\{x_k\}$ en $\mathbf{X}$ que satisface

$$f(x_k) \to \text{ínf}_{\mathbf{X}}(f) \quad y \quad \nabla f(x_k) \to 0.$$

Demostración. Supongamos que $\lambda = 1$ (en caso contrario basta considerar la norma $\|\cdot\|/\lambda$). Definamos la función $g_\varepsilon \colon \mathbf{X} \to \mathbb{R} \cup \{+\infty\} \colon x \mapsto f(x) + \varepsilon\|x - x_0\|$. Es fácil mostrar que g_ε es convexa, s.c.i. y coerciva, por lo que el Teorema 3.1 implica que el conjunto $\operatorname{sol}(G_\varepsilon)$ es no vacío, donde

$$(G_\varepsilon) \qquad \operatorname{val}(G_\varepsilon) = \inf_{x \in \mathbf{X}} g_\varepsilon(x).$$

Además, como

$$\operatorname{sol}(G_\varepsilon) = \{x \in \mathbf{X} \,|\, g_\varepsilon(x) \leq \operatorname{val}(G_\varepsilon)\} = \Gamma_{\operatorname{val}(G_\varepsilon)}(g_\varepsilon),$$

se tiene que $\operatorname{sol}(G_\varepsilon)$ es convexo, cerrado y acotado. De ese modo, como $f + \delta_{\operatorname{sol}(G_\varepsilon)}$ es propia, s.c.i. y coerciva, el Teorema 3.1 implica que existe $x_\varepsilon \in \operatorname{sol}(P_\varepsilon)$, donde

$$(P_\varepsilon) \qquad \operatorname{val}(P_\varepsilon) = \inf_{x \in \operatorname{sol}(G_\varepsilon)} f(x).$$

Luego, como $x \in \operatorname{sol}(G_\varepsilon)$, se tiene que $g_\varepsilon(x_\varepsilon) \leq g_\varepsilon(x)$ para todo $x \in \mathbf{X}$, lo que implica

$$(4.3) \qquad \inf_{x \in \mathbf{X}} f(x) + \varepsilon\|x_\varepsilon - x_0\| \leq f(x_\varepsilon) + \varepsilon\|x_\varepsilon - x_0\| = g_\varepsilon(x_\varepsilon) \leq g_\varepsilon(x_0) = f(x_0) \leq \inf_{x \in \mathbf{X}} f(x) + \varepsilon,$$

donde la última desigualdad corresponde a (4.2). De la cadena de desigualdades anteriores se concluye $f(x_\varepsilon) \leq f(x_0)$ y $\|x_\varepsilon - x_0\| \leq 1$.

Sea $x \in \mathbf{X} \setminus \{x_\varepsilon\}$. Si $x \in \mathbf{X} \setminus \operatorname{sol}(G_\varepsilon)$, entonces $g_\varepsilon(x_\varepsilon) < g_\varepsilon(x)$, lo que implica

$$f(x_\varepsilon) < f(x) + \varepsilon(\|x - x_0\| - \|x_\varepsilon - x_0\|) \leq f(x) + \varepsilon\|x - x_\varepsilon\|.$$

Si por el contrario $x \in \operatorname{sol}(G_\varepsilon) \setminus \{x_\varepsilon\}$, entonces, como $x_\varepsilon \in \operatorname{sol}(P_\varepsilon)$,

$$f(x_\varepsilon) \leq f(x) < f(x) + \varepsilon\|x - x_\varepsilon\|.$$

Concluimos que, para todo $x \in \mathbf{X} \setminus \{x_\varepsilon\}$ se tiene $f(x_\varepsilon) < f(x) + \varepsilon\|x - x_\varepsilon\|$.

Para las últimas afirmaciones, supongamos que f es Gâteaux diferenciable y sea $d \in \mathbf{X}$ con $\|d\| = 1$. Por la parte (iii),

$$(\forall t \neq 0) \quad \frac{f(x_\varepsilon) - f(x_\varepsilon + td)}{t} \leq \varepsilon\|d\| = \varepsilon,$$

de donde, haciendo $t \to 0$, tenemos que

$$-\langle \nabla f(x_\varepsilon), d \rangle \leq \varepsilon.$$

Tomando $d = -\nabla f(x_\varepsilon)/\|\nabla f(x_\varepsilon)\|$, se concluye $\|\nabla f(x_\varepsilon)\| \leq \varepsilon$. Finalmente, para cada $k \in \mathbb{N}$ tomemos $y_k \in \frac{1}{k} - \arg\min_{\mathbf{X}}(f)$. Luego, basta aplicar el resultado recientemente probado para obtener la existencia de $x_k \in \mathbf{X}$ que satisface

$$f(x_k) \leq f(y_k) \leq \inf_{\mathbf{X}}(f) + \frac{1}{k} \quad \text{y} \quad \|\nabla f(x_k)\| \leq \frac{1}{k}.$$

□

4.5. Métodos de descenso

El principio Variacional de Ekeland provee de forma abstracta la existencia de una sucesión minimizante $\{x_k\}$ tal que $\nabla f(x_k) \to 0$. Veremos ahora dos métodos constructivos que permiten determinar una tal sucesión usando la información entregada por los datos del problema. En lo que sigue supondremos que $\mathbf{X}$ tiene la estructura de espacio de Hilbert con un producto interno $\langle \cdot, \cdot \rangle$. En este contexto, el siguiente lema será de utilidad para la convergencia en espacios de Hilbert de varios métodos en este curso.

Lema 4.1 (Opial)**.** *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert. Sea $\{x_k\}$ una sucesión en $\mathbf{X}$ y $\mathbf{S} \subset \mathbf{X}$ un conjunto no vacío. Supongamos que:*

(a) para todo $x \in \mathbf{S}$, la sucesión $\{\|x_k - x\|\}$ converge;

(b) todo punto de acumulación débil de $\{x_k\}$ está en $\mathbf{S}$.

Entonces, existe $\bar{x} \in \mathbf{S}$ tal que $x_k \rightharpoonup \bar{x}$ cuando $k \to \infty$.

Demostración. Sean x e y dos puntos de acumulación débiles de $\{x_k\}$, digamos $x_{k_n} \rightharpoonup x$ y $x_{k_m} \rightharpoonup y$. Estos existen pues $\{x_k\}$ es una sucesión acotada por (a). De (b) se obtiene que x e y están en $\mathbf{S}$. Además, como se cumple

$$\|x_k - x\|^2 - \|x_k - y\|^2 = -\langle x - y, 2x_k - x - y \rangle, \qquad \forall k \in \mathbb{N},$$

y el lado izquierdo converge a un límite ℓ. Tomando en particular las subsucesiones $\{x_{k_n}\}$ y $\{x_{k_m}\}$ al lado derecho se concluye $\ell = \|x - y\|^2 = -\|x - y\|^2$, de donde obtenemos que $x = y$. □

4.5.1. Método del Gradiente

El primer método que estudiaremos se basa en una iteración del tipo

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad \forall k \in \mathbb{N} \tag{4.4}$$

que parte desde $x_0 \in \mathbf{X}$ arbitrario y donde $\alpha_k > 0$ para cada $k \in \mathbb{N}$. Notemos que, gracias a la Regla de Fermat, si en alguna iteración tenemos que x_k es un mínimo de f entonces

$$x_l = x_k, \quad \forall l \geq k$$

y por lo tanto el método se detiene una vez que se llega a un óptimo.

Para estudiar la convergencia del método del gradiente, necesitamos algunas propiedades de funciones convexas diferenciable con gradiente Lipschitz continuo.

Lema 4.2 (Lema de máximo descenso)**.** *Sean $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y $f \colon \mathbf{X} \to \mathbb{R}$ una función Gâteux diferenciable en $\mathbf{X}$ tal que ∇f es L-Lipschitz continuo en $\mathbf{X}$, es decir,*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbf{X}.$$

Entonces se cumple

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2, \quad \forall x, y \in \mathbf{X}. \tag{4.5}$$

Además, si f es convexa, se tiene

$$f(y) \leq f(z) + \langle \nabla f(x), y - z \rangle + \frac{L}{2}\|y - x\|^2, \quad \forall x, y, z \in \mathbf{X}. \tag{4.6}$$

Demostración. Sean x e y en $\mathbf{X}$. Definamos, para todo $t \in [0,1]$ la función $\phi(t) = f(x+t(y-x))$. Usando la propiedad Lipschitzianidad de ∇f y la desigualdad de Cauchy-Schwartz, se tiene

$$\begin{aligned} f(y)-f(x) &= \int_0^1 \phi'(t)dt \\ &= \int_0^1 \langle \nabla f(x+t(y-x)), y-x\rangle dt \\ &= \langle \nabla f(x), y-x\rangle + \int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x\rangle dt \\ &\leq \langle \nabla f(x), y-x\rangle + \int_0^1 \|\nabla f(x+t(y-x)) - \nabla f(x)\| \, \|y-x\| dt \\ &\leq \langle \nabla f(x), y-x\rangle + L\|y-x\|^2 \int_0^1 t dt \\ &= \langle \nabla f(x), y-x\rangle + \frac{L}{2}\|y-x\|^2, \end{aligned}$$

de donde se obtiene la primera desigualdad (4.5). Para la segunda, dado $z \in \mathbf{X}$, gracias al Teorema 4.1, como f es convexa y Gâteaux diferenciable en $\mathbf{X}$, se tiene que

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

Luego, basta sumar esta desigualdad a (4.5) para obtener (4.6). □

Ahora podemos estudiar la convergencia del método del gradiente.

Teorema 4.6. *Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ es un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R}$ una función convexa Gâteaux diferenciable en $\mathbf{X}$ y tal que $\arg\min_{\mathbf{X}}(f) \neq \emptyset$. Supongamos que ∇f es L-Lipschitz continuo en $\mathbf{X}$. Considere la sucesión $\{x_k\}_{k\in\mathbb{N}}$ generada por (4.4) partiendo desde $x_0 \in \mathbf{X}$, con*

$$0 < \alpha_{min} \leq \alpha_k \leq \alpha_{max} < \frac{2}{L}, \qquad \forall k \in \mathbb{N}.$$

Entonces $\exists x_\infty \in \arg\min_{\mathbf{X}}(f)$ tal que $x_k \rightharpoonup x_\infty$ cuando $k \to \infty$.

Demostración. Dividamos la demostraciones en partes:

1. Veamos primero que la sucesión $\{f(x_k)\}$ es decreciente y convergente. Sea $k \in \mathbb{N}$. Tomando $y = x_{k+1}$ y $x = x_k$ en (4.6), y usando (4.4), se tiene

$$f(x_{k+1}) \leq f(z) + \langle \nabla f(x_k), x_k - z\rangle - \alpha_k\left(1 - \frac{\alpha_k L}{2}\right)\|\nabla f(x_k)\|^2, \quad \forall z \in \mathbf{X}. \tag{4.7}$$

   En particular, tomando $z = x_k$ y usando $\alpha_{max} < 2/L$, se obtiene el decrecimiento de la sucesión $\{f(x_k)\}$, y por lo tanto, la convergencia de esta sucesión, pues $\inf_{\mathbf{X}}(f) > -\infty$.

2. Evaluando $z = x_k$ en (4.7) y sumando sobre k desde 0 a n, se deduce de la propiedad telescópica y de las hipótesis sobre los α_k

$$\alpha_{min}\left(1 - \frac{\alpha_{max}L}{2}\right)\sum_{k=0}^{n}\|\nabla f(x_k)\|^2 \leq f(x_0) - f(x_{n+1}), \qquad \forall n \in \mathbb{N}$$

de donde podemos inferir que

$$\sum_{k=0}^{\infty} \|\nabla f(x_k)\|^2 < +\infty \quad \text{y por lo tanto } \nabla f(x_k) \to 0 \text{ para la topología fuerte si } k \to +\infty.$$

3. Ahora tomando $z = \bar{x} \in \arg\min_{\mathbf{X}}(f)$ en (4.7), deducimos

$$\begin{aligned}\|x_{k+1} - \bar{x}\|^2 - \|x_k - \bar{x}\|^2 - \alpha_k^2 \|\nabla f(x_k)\|^2 &= 2\alpha_k \langle \nabla f(x_k), \bar{x} - x_k \rangle \\ &\leq 2\alpha_k \left(f(\bar{x}) - f(x_{k+1}) - \alpha_k \left(1 - \frac{\alpha_k L}{2}\right) \|\nabla f(x_k)\|^2 \right) \\ &\leq -\alpha_k^2 (2 - \alpha_k L) \|\nabla f(x_k)\|^2,\end{aligned}$$

de donde obtenemos $\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 + s_k$, donde

$$s_k := \alpha_{max}^3 L \|\nabla f(x_k)\|^2.$$

4. Afirmamos que la sucesión $\{\|x_k - \bar{x}\|\}$ es convergente. En efecto, sumando a ambos lados de la última desigualdad el término $\sum_{l=k+1}^{\infty} s_k$ obtenemos

$$\theta_{k+1} := \|x_{k+1} - \bar{x}\|^2 + \sum_{l=k+1}^{\infty} s_k \leq \|x_k - \bar{x}\|^2 + \sum_{l=k}^{\infty} s_k =: \theta_k$$

Luego la sucesión $\{\theta_k\}$ es decreciente y y todos sus términos son no negativos. Luego, $\{\theta_k\}$ es convergente, pero como $\sum_{l=k}^{\infty} s_k \to 0$ si $k \to +\infty$ (pues la serie converge), se tiene que la sucesión $\{\|x_k - \bar{x}\|\}$ también converge.

5. Para concluir usamos el Lema 4.1. Tomemos un punto de acumulación débil de $\{x_k\}$, digamos $x_{k_n} \rightharpoonup y \in \mathbf{X}$ si $n \to +\infty$. Tomando $z = \bar{x}$ en (4.7) y usando la semicontinuidad inferior de f en la topología débil (por convexidad) deducimos que

$$f(y) \leq \liminf_{n\to\infty} f(x_{k_n+1}) \leq f(\bar{x}) + \liminf_{n\to\infty} \left(\langle \nabla f(x_{k_n}), x_{k_n} - \bar{x} \rangle - \alpha_{k_n} \left(1 - \frac{\alpha_{k_n} L}{2}\right) \|\nabla f(x_{k_n})\|^2 \right).$$

Sabemos que $\|\nabla f(x_{k_n})\| \to 0$ si $n \to +\infty$ y que además $\{x_{k_n}\}$ está acotada (pues converge débilmente). Luego, debido a las condiciones sobre $\{\alpha_k\}$, el limite inferior de la derecha es nulo y por lo tanto debemos tener que $f(y) \leq f(\bar{x})$. Lo que implica a su vez que $y \in \arg\min_{\mathbf{X}}(f)$ y en consecuencia el resultado final de convergencia se deduce del Lema 4.1 con $\mathbf{S} = \arg\min_{\mathbf{X}}(f)$.

□

Aplicación al problema cuadrático

Estudiaremos ahora una versión especialidad del método del gradiente para minimizar funciones cuadráticas del tipo

$$f(x) = \frac{1}{2} x^\top A x - b^\top x + c, \quad \forall x \in \mathbb{R}^n,$$

donde $A \in \mathbb{S}^n_{++}(\mathbb{R})$, $b \in \mathbb{R}^n$ y $c \in \mathbb{R}$. Sabemos, por la Proposición 4.2 que existe un único $\bar{x}$ que minimiza esta función. Una forma de aproximar $\bar{x}$ es utilizando el método del gradiente, cuya construcción iterativa está dada por (4.4). En este caso estudiaremos la velocidad de convergencia del método con α_k siendo el único real positivo que minimiza sobre $\mathbb{R}$ la aplicación

$$\alpha \mapsto f(x_k - \alpha \nabla f(x_k)).$$

Veremos que la velocidad de convergencia de $\{x_k\}$ a $\bar{x}$ depende de un real asociado a la matriz A llamado *condicionamiento*, el cual está dado por

$$\kappa(A) := \frac{\lambda_1}{\lambda_n}$$

donde $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$ los valores propios de A en orden decreciente.

Antes de continuar, notemos que la aplicación

$$\alpha \mapsto f(x_k - \alpha \nabla f(x_k))$$

es estrictamente convexa y diferenciable, luego, usando la Regla de Fermat, α_k puede ser calculado explícitamente. De hecho, tenemos que

$$\alpha_k = \frac{\|Ax_k - b\|^2}{(Ax_k - b)^\top A(Ax_k - b)} = \frac{\|\nabla f(x_k)\|^2}{\nabla f(x_k)^\top A \nabla f(x_k)}, \qquad \forall k \in \mathbb{N}. \tag{4.8}$$

Teorema 4.7. *Sean $A \in \mathbb{S}^n_{++}(\mathbb{R})$, $b \in \mathbb{R}^n$ y $c \in \mathbb{R}$. Considere la función $f : \mathbb{R}^n \to \mathbb{R}$ dada por*

$$f(x) = \frac{1}{2} x^\top A x - b^\top x + c, \quad \forall x \in \mathbb{R}^n,$$

y sea $\bar{x}$ su mínimo. Considere la sucesión $\{x_k\}_{k\in\mathbb{N}}$ generada por (4.4) partiendo desde $x_0 \in \mathbb{R}^n$ arbitrario y con α_k dado por (4.8). Luego se tienen las estimaciones

$$(i) \qquad f(x_k) - f(\bar{x}) \leq [f(x_0) - f(\bar{x})] \left[\frac{\kappa(A) - 1}{\kappa(A) + 1}\right]^{2k}$$

$$(ii) \qquad \|x_k - \bar{x}\|_A \leq \|x_0 - \bar{x}\|_A \left[\frac{\kappa(A) - 1}{\kappa(A) + 1}\right]^{k}$$

$$(iii) \qquad \|x_k - \bar{x}\| \leq \left[\frac{2(f(x_0) - f(\bar{x}))}{\lambda_n}\right]^{\frac{1}{2}} \left[\frac{\kappa(A) - 1}{\kappa(A) + 1}\right]^{k},$$

donde $\|\cdot\|_A : x \mapsto \sqrt{x^\top A x}$ define una norma en $\mathbb{R}^n$.

Demostración. Para todo $k \in \mathbb{N}$, definamos $g_k := \nabla f(x_k) = Ax_k - b$, de donde

$$\alpha_k = \frac{\|g_k\|^2}{g_k{}^\top A g_k} \quad \text{y} \quad x_{k+1} = x_k - \alpha_k g_k, \quad \forall k \in \mathbb{N}.$$

Entonces, sigue que

$$\begin{aligned} f(x_{k+1}) &= \frac{1}{2}{x_{k+1}}^{\top} A x_{k+1} - b^{\top} x_{k+1} + c \\ &= \frac{1}{2}{x_k}^{\top} A x_k - \alpha_k (A x_k)^{\top} g_k + \frac{\alpha_k^2}{2}{g_k}^{\top} A g_k - b^{\top} x_k + \alpha_k b^{\top} g_k + c \\ &= f(x_k) - \alpha_k (A x_k - b)^{\top} g_k + \frac{\alpha_k^2}{2}{g_k}^{\top} A g_k \end{aligned}$$

En particular, tenemos que

$$f(x_{k+1}) = f(x_k) - \frac{\|g_k\|^4}{2{g_k}^{\top} A g_k}, \quad \forall k \in \mathbb{N}. \tag{4.9}$$

Además, como la única solución óptima del problema está dada por $\bar{x} = A^{-1}b$, deducimos que $\inf_{\mathbf{X}}(f) = f(\bar{x}) = -\frac{1}{2}b^{\top}A^{-1}b + c$ y más aún

$$\frac{1}{2}(A^{-1}g_k)^{\top} g_k = \frac{1}{2}(x_k - A^{-1}b)^{\top}(Ax_k - b) = \frac{1}{2}{x_k}^{\top} A x_k - b^{\top} x_k + \frac{1}{2}b^{\top}A^{-1}b = f(x_k) - f(\bar{x}), \quad \forall k \in \mathbb{N}.$$

Luego, de (4.9) se deduce

$$f(x_{k+1}) - f(\bar{x}) = [f(x_k) - f(\bar{x})]\left(1 - \frac{\|g_k\|^4}{\left({g_k}^{\top} A g_k\right)\left({g_k}^{\top} A^{-1} g_k\right)}\right), \quad \forall k \in \mathbb{N}.$$

Entonces, como la desigualdad de Kantorovich (ver Ejercicio 4) asegura que

$$\left(x^{\top} A x\right)\left(x^{\top} A^{-1} x\right) \leq \frac{(\kappa(A)+1)^2}{4\kappa(A)}\|x\|^4, \quad \forall x \in \mathbb{R}^n,$$

conluimos

$$f(x_{k+1}) - f(\bar{x}) \leq [f(x_k) - f(\bar{x})]\left(1 - \frac{4\kappa(A)}{(\kappa(A)+1)^2}\right) = [f(x_k) - f(\bar{x})]\left(\frac{\kappa(A)-1}{\kappa(A)+1}\right)^2, \quad \forall k \in \mathbb{N}$$

y la primera desigualdad se obtiene usando inducción. Además, dado que

$$2(f(x_k) - f(\bar{x})) = {g_k}^{\top} A^{-1} g_k = (x_k - \bar{x})^{\top} A (x_k - \bar{x}) = \|x_k - \bar{x}\|_A^2, \quad \forall k \in \mathbb{N},$$

usando la primera desigualdad se deduce directamente la segunda. Finalmente, la última desigualdad se deduce de la segunda, ya que gracias a la Proposición 4.2 tenemos que

$$\|x_k - \bar{x}\|_A^2 \geq \lambda_n \|x_k - \bar{x}\|^2, \quad \forall k \in \mathbb{N}.$$

□

4.5.2. Método del Gradiente conjugado

Ahora veremos un método en el contexto $\mathbf{X} = \mathbb{R}^n$ cuya principal característica es que encuentra en una cantidad finita de iteraciones el óptimo de una función cuadrática estrictamente convexa. La idea principal de este algoritmo se basa en el hecho que para una iteración del tipo

$$x_{k+1} = x_k + \alpha_k d_k, \quad \forall k \in \mathbb{N}, \tag{4.10}$$

con $d_k \in \mathbb{R}^n$ cualquiera, si α_k se escoge como un real que minimiza sobre $\mathbb{R}$ la función convexa

$$\alpha \mapsto f(x_k + \alpha d_k),$$

por la Regla de Fermat se tendrá que $\nabla f(x_{k+1})$ y d_k son ortogonales. El método consisten entonces en escoger los d_k de forma tal que $\nabla f(x_{k+1})$ sea ortogonal no solo a d_k, si no que también a $d_0, \ldots, d_{k-1}$. De esta forma, al cabo de n iteraciones se deberá tener forzosamente que $\nabla f(x_n) = 0$ y que por lo tanto x_n es un mínimo de la función.

El nombre del método viene del hecho que dos vectores $x, y \in \mathbb{R}^n$ se dicen *conjugados* con respecto a $A \in \mathbb{S}^n_{++}(\mathbb{R})$ si $x^\top A y = 0$. Notemos que, para cualquier $k \in \{1, \ldots, n-1\}$, si $v_1, \ldots, v_{k+1} \in \mathbb{R}^n$ son vectores no nulos conjugados con respecto a A, entonces $\{v_1, \ldots, v_{k+1}\}$ es una familia linealmente independiente. En efecto, si esto no fuese así, podemos asumir sin perdida de generalidad que v_{k+1} se puede escribir como combinación lineal de $v_1, \ldots, v_k$, es decir, existen $\xi_1, \ldots, \xi_k \in \mathbb{R}$ tales que

$$v_{k+1} = \sum_{k=1}^{k} \xi_i v_i \quad \text{y por lo tanto} \quad v_{k+1}^\top A v_{k+1} = v_{k+1}^\top \left(\sum_{i=1}^{k} \xi_i A v_i \right) = \sum_{i=1}^{k} \xi_i v_{k+1}^\top A v_i = 0.$$

Dado que $A \in \mathbb{S}^n_{++}(\mathbb{R})$ y $v_{k+1} \neq 0$ llegamos a una contradicción. Notemos que esto implica que una colección de vectores no nulos conjugados con respecto a A no puede contener más de n vectores.

El método del gradiente conjugado consiste en utilizar las direcciones d_k dadas por la fórmula

$$d_k = \begin{cases} -g_0 & \text{si } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 1, \end{cases} \tag{4.11}$$

donde denotamos $g_k := \nabla f(x_k) = Ax_k - b$ para todo $k \in \mathbb{N}$ y β_k es un parámetro dado por la relación de conjugación entre d_k y d_{k-1}. Efectivamente, no es difícil ver que ${d_k}^\top A d_{k-1} = 0$ si y sólo si

$$\beta_k = \frac{(Ax_k - b)^\top A d_{k-1}}{{d_{k-1}}^\top A d_{k-1}} = \frac{{g_k}^\top A d_{k-1}}{{d_{k-1}}^\top A d_{k-1}}, \qquad \forall k \in \mathbb{N} \setminus \{0\}. \tag{4.12}$$

Más aún, si el paso se escoge de forma óptima, gracias a la Regla de Fermat tenemos que

$$\alpha_k = -\frac{(Ax_k - b)^\top d_k}{{d_k}^\top A d_k} = -\frac{{g_k}^\top d_k}{{d_k}^\top A d_k}, \qquad \forall k \in \mathbb{N}. \tag{4.13}$$

Ahora veremos que el método converge en una cantidad finita de pasos

Teorema 4.8. *Sean $A \in \mathbb{S}^n_{++}(\mathbb{R})$, $b \in \mathbb{R}^n$ y $c \in \mathbb{R}$. Considere la función $f : \mathbb{R}^n \to \mathbb{R}$ dada por*

$$f(x) = \frac{1}{2} x^\top A x - b^\top x + c, \quad \forall x \in \mathbb{R}^n.$$

La sucesión $\{x_k\}_{k \in \mathbb{N}}$ generada por (4.10) partiendo desde $x_0 \in \mathbb{R}^n$ arbitrario, con d_k dado por (4.11), β_k dado por (4.12) y α_k dado por (4.13), converge en a lo más n pasos.

Demostración. Procederemos por inducción y probaremos que, para todo $k \in \{1,\dots,n\}$, si $d_{k-1} \neq 0$ entonces

$$g_k{}^\top d_i = d_k{}^\top A d_i = 0, \quad \forall i \in \{0,\dots,k-1\}. \tag{4.14}$$

De este modo, si para algún $k \in \{1,\dots,n\}$, tuviesemos $d_k = 0$, entonces $0 = d_k = -g_k + \beta_k d_{k-1}$, de donde g_k es colineal con $d_{k-1} \neq 0$ y al mismo tiempo $g_k{}^\top d_{k-1} = 0$, por lo que $g_k = 0$ y x_k es solución. Además, el algoritmo terminará en a lo más n pasos, ya que a cada iteración k genera un d_k que es conjugado con respecto a A a $d_0,\dots,d_{k-1}$, lo cual se puede hacer a lo más n veces en $\mathbb{R}^n$.

Para $k = 1$, si $d_0 = -g_0 \neq 0$, de (4.10) y (4.13) deducimos

$$g_1{}^\top d_0 = (Ax_1 - b)^\top d_0 = g_0{}^\top d_0 + \alpha_0 d_0{}^\top A d_0 = 0.$$

Además, de (4.11) y (4.12) obtenemos

$$d_1{}^\top A d_0 = -g_1{}^\top A d_0 + \beta_1 d_0{}^\top A d_0 = 0.$$

Ahora supongamos que para $k \in \{1,\dots,n-1\}$, si se tiene $d_{k-1} \neq 0$, entonces se cumple (4.14). Supongamos que $d_k \neq 0$ y tomemos $i \in \{0,\dots,k\}$. Si $i = k$ entonces (4.10) y (4.13) implican que

$$g_{k+1}{}^\top d_k = (Ax_{k+1} - b)^\top d_k = g_k{}^\top d_k + \alpha_k d_k{}^\top A d_k = 0.$$

Además, de (4.11) y (4.12) vemos que

$$d_{k+1}{}^\top A d_k = -g_{k+1}{}^\top A d_k + \beta_{k+1} d_k{}^\top A d_k = 0.$$

Ahora, si $i < k$, (4.10) implica que $g_{k+1} = g_k + \alpha_k A d_k$ de donde

$$g_{k+1}{}^\top d_i = g_k{}^\top d_i + \alpha_k d_k{}^\top A d_i = 0,$$

pues ambos términos son nulos por la hipótesis de inducción. Por otra parte, combinando el hecho que $g_{i+1} = g_i + \alpha_i A d_i$ con (4.11) deducimos que

$$A d_i = \frac{1}{\alpha_i}(g_{i+1} - g_i) = \begin{cases} \dfrac{(\beta_1 + 1)d_0 - d_1}{\alpha_0} & \text{si } i = 0, \\ \dfrac{1}{\alpha_i}(\beta_{i+1} d_i - d_{i+1} - (\beta_i d_{i-1} - d_i)) & \text{si } i \in \{1,\dots,k-1\}, \end{cases}$$

de donde obtenemos,

$$d_{k+1}{}^\top A d_i = -g_{k+1}{}^\top A d_i + \beta_{k+1} d_k{}^\top A d_i = 0, \quad \forall i \in \{1,\dots,k-1\},$$

pues el segundo término es nulo por hipótesis de inducción y el primero es nulo pues acabamos de probar que $g_{k+1}{}^\top d_i = 0$ para todo $i \in \{1,\dots,k\}$ y $A d_i$ es una combinación lineal de d_{i-1}, d_i, d_{i+1} (y $A d_0$ es combinación lineal de d_0, d_1). Esto concluye la demostración. □

En el caso que n sea muy grande (por ejemplo $n \geq 10^3$), realizar las n iteraciones del método del gradiente conjugado puede ser muy costoso. Por esta razón, es interesante saber lo preciso que se vuelve el método al cabo de algunas iteraciones. El siguiente resultado provee una estimación de este error y entrega una cota para la tasa de convergencia del método.

Teorema 4.9. *Sean $A \in \mathbb{S}^n_{++}(\mathbb{R})$, $b \in \mathbb{R}^n$ y $c \in \mathbb{R}$. Considere la función $f : \mathbb{R}^n \to \mathbb{R}$ dada por*

$$f(x) = \frac{1}{2}x^\top Ax - b^\top x + c$$

y sea $\bar{x}$ su mínimo. Considere la sucesión $\{x_k\}$ generada por (4.10) partiendo desde $x_0 \in \mathbf{X}$ con d_k dado por (4.11), β_k dado por (4.12) y α_k dado por (4.13). Luego

$$\|x_{k+1} - \bar{x}\|_A \leq 2\|x_0 - \bar{x}\|_A \left[\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1}\right]^k, \qquad \forall k \in \mathbb{N},$$

donde $\|x\|_A = \sqrt{x^\top Ax}$ para cada $x \in \mathbb{R}^n$.

Demostración. Ejercicio de ayudantía. □

Comparación de los métodos

En el contexto de problemas cuadráticos estrictamente convexo, es decir $A \in \mathbb{S}^n_{++}(\mathbb{R})$, la tasa de convergencia del método gradiente conjugado es mejor que la tasa del método del gradiente, ya que

$$0 \leq \frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \leq \frac{\kappa(A) - 1}{\kappa(A) + 1} < 1,$$

con igualdad solo para el caso $\kappa(A) = 1$. Esto implica que en general se tiene que el método gradiente conjugado convergerá más rápido que el método del gradiente.

4.5.3. Método de Newton-Raphson

El método del gradiente considera información sólo de primer orden, lo cual, dependiendo de la función a minimizar, puede generar algoritmos que convergen muy lentos debido a un efecto de *zig-zag* como lo explica el siguiente ejemplo.

Ejemplo 4.5.1. *Sea $\delta > 1$ y $f\colon \mathbb{R}^2 \to \mathbb{R}$ definido por*

$$f(x,y) = \frac{1}{2}x^2 + \frac{\delta}{2}y^2, \quad \forall x, y \in \mathbb{R}.$$

Se tiene $\nabla f(x,y) = (x, \delta y)$ y que por lo tanto es es $\delta-$Lipschitz continuo. El método del gradiente se escribe en este caso como

$$\begin{cases} x_{k+1} &= x_k(1-\alpha) \\ y_{k+1} &= y_k(1-\alpha\delta), \end{cases}$$

donde $\alpha < 2/\delta$. En particular, si $\alpha = 0{,}1$ y $\delta = 18$, la condición de los parámetros para la convergencia se satisfacen y el método se reduce a

$$\begin{cases} x_{k+1} &= 0{,}9 * x_k \\ y_{k+1} &= -0{,}8 * y_k, \end{cases}$$

lo que hace zig-zaguear *a las iteraciones si $y_k \neq 0$ y la convergencia es cada vez más lenta si δ es mayor y α es más pequeño. El siguiente Cuadro muestra algunas iteraciones del método del gradiente con $(x_0, y_0) = (1,1)$:*

k	x_k	y_k	k	x_k	y_k	k	x_k	y_k
0	1.0000	1.0000	7	0.4783	-0.2097	14	0.2288	0.0440
1	0.9000	-0.8000	8	0.4305	0.1678	15	0.2059	-0.0352
2	0.8100	0.6400	9	0.3874	-0.1342	16	0.1853	0.0281
3	0.7290	-0.5120	10	0.3487	0.1074	17	0.1668	-0.0225
4	0.6561	0.4096	11	0.3138	-0.0859	18	0.1501	0.0180
5	0.5905	-0.3277	12	0.2824	0.0687	19	0.1351	-0.0144
6	0.5314	0.2621	13	0.2542	-0.0550	20	0.1216	0.0115

Cuadro 4.1: Iteraciones del método del gradiente

Estudiaremos ahora otro método, llamado *Newton-Raphson*, que involucra la curvatura de la función a minimizar, lo que permite superar estos efectos que reducen la velocidad de convergencia. La idea principal es, conocida la iteración $x_k \in \mathbf{X} := \mathbb{R}^n$, minimizar la aproximación de Taylor de segundo orden de f en torno a x_k

$$f_k(x) = f(x_k) + \nabla f(x_k)^\top (x - x_k) + \frac{1}{2}(x - x_k)^\top \nabla^2 f(x_k)(x - x_k)$$

para encontrar x_{k+1}, donde $\nabla^2 f(x)$ es la matrix hessiana de f en x. Usando la regla de Fermat, lo anterior se traduce a resolver la ecuación para x_{k+1}

$$0 = \nabla f(x_k) + \nabla^2 f(x_k)(x_{k+1} - x_k),$$

que, en el caso en que $\nabla^2 f(x_k)$ sea invertible, se reduce a

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k), \quad \forall k \in \mathbb{N}. \tag{4.15}$$

Ahora veamos como cambia la eficiencia del método de Newton-Raphson en comparación al método del gradiente para el ejemplo anterior.

Ejemplo 4.5.2. *Retomemos el Ejemplo 4.5.1. Recordemos que $f(x,y) = \frac{1}{2}x^2 + \frac{\delta}{2}y^2$. Es claro que el único mínimo de esta función es $(\bar{x}, \bar{y}) = (0,0)$. Además, tenemos que, y*

$$\nabla f(x,y) = \begin{pmatrix} x \\ \delta y \end{pmatrix} \quad y \quad \nabla^2 f(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & \delta \end{pmatrix}.$$

Luego, dado $(x_0, y_0) \in \mathbb{R}^2$, la primera iteración es

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\delta} \end{pmatrix} \begin{pmatrix} x_0 \\ \delta y_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

es decir, el método de Newton-Raphson encuentra el mínimo en una sola iteración.

Observación 4.1. *Más generalmente, el método de Newton-Raphson es utilizado para la resolución de ecuaciones no lineales del tipo $F(x) = 0$, donde $F : \mathbb{R}^n \to \mathbb{R}^n$ es una función Fréchet diferenciable. En este contexto, dado x_k, el método busca x_{k+1} resolviendo la aproximación de primer orden*

$$0 = F(x_k) + J_F(x_k)(x_{k+1} - x_k),$$

donde, si la matriz Jacobiana $J_F(x_k)$ es invertible, se reduce a

$$x_{k+1} = x_k - J_F(x_k)^{-1} F(x_k).$$

Recuerdo : Matriz Jacobiana

Una función vectorial $F:\mathbb{R}^n\to\mathbb{R}^m$ se dice Fréchet diferenciable en $x\in\mathbb{R}^n$ si existe una matriz $M\in\mathbb{M}_{m\times n}(\mathbb{R})$ tal que

$$\lim_{h\to 0}\frac{|F(x+h)-F(x)-Mh|}{|h|}=0.$$

La matriz M se denota $J_F(x)$, se conoce como la Matriz Jacobiana de F y viene dada por:

$$J_F(x)=\begin{pmatrix}\partial_{x_1}F_1(x) & \dots & \partial_{x_n}F_1(x)\\ \vdots & \ddots & \vdots\\ \partial_{x_1}F_m(x) & \dots & \partial_{x_n}F_m(x)\end{pmatrix}$$

donde $F(x)=(F_1(x),\dots,F_m(x))$ para todo $x\in\mathbb{R}^n$

Ahora estudiaremos la convergencia del método de Newton-Raphson. Cabe destacar que el teorema de convergencia que mostraremos ahora se diferencia de los teoremas estudiados para los métodos del Gradiente y Gradiente conjugado en que la elección del punto inicial juega un rol importante.

Teorema 4.10. *Sea $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ una función propia, convexa y dos veces Gâteaux diferenciable en $dom(f)$, el cual asumimos abierto de $\mathbb{R}^n$. Supongamos que existe $\bar{x}\in\arg\min_{\mathbb{R}^n}(f)$ tal que $\nabla^2 f(\bar{x})\in\mathbb{S}^n_{++}(\mathbb{R})$, y que además $\nabla^2 f$ es localmente Lipschitz continua en torno a $\bar{x}$, es decir, para algún $r>0$ existe $L>0$ tal que*

$$\|\nabla^2 f(x)-\nabla^2 f(y)\|\le L|x-y|,\quad \forall x,y\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},r),$$

donde $\|M\|=\sup_{\|x\|=1}\|Mx\|=\sqrt{\lambda_{\max}(M^\top M)}$ para cualquier $M\in\mathbb{M}_{n\times n}(\mathbb{R})$. Entonces, existe $\rho>0$ para el cual se tiene que si $x_0\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$, la secuencia $\{x_k\}$ generada por (4.15) converge a $\bar{x}$ y satisface

$$\lim_{k\to\infty}\frac{|x_{k+1}-\bar{x}|}{|x_k-\bar{x}|}=0,\quad y\quad \limsup_{k\to\infty}\frac{|x_{k+1}-\bar{x}|}{|x_k-\bar{x}|^2}<\infty.$$

Demostración. Para todo $x\in\mathrm{dom}(f)$ denotemos por λ_x al menor valor propio de $\nabla^2 f(x)$. Como $\nabla^2 f(\bar{x})\in\mathbb{S}^n_{++}(\mathbb{R})$, de la Proposición 4.2 se tiene

$$y^\top\nabla^2 f(\bar{x})y\ge\lambda_{\bar{x}}|y|^2,\quad\forall y\in\mathbb{R}^n,$$

donde $\lambda_{\bar{x}}>0$. Para todo $x\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},r)$ e $y\in\mathbb{R}^n$, usando la propiedad Lipschitz de $\nabla^2 f$ se tiene

$$\begin{aligned}y^\top\nabla^2 f(x)y&=y^\top\nabla^2 f(\bar{x})y+y^\top(\nabla^2 f(x)-\nabla^2 f(\bar{x}))y\\&\ge\lambda_{\bar{x}}|y|^2-\|\nabla^2 f(x)-\nabla^2 f(\bar{x})\||y|^2\\&\ge(\lambda_{\bar{x}}-L|x-\bar{x}|)|y|^2.\end{aligned}$$

Luego, definiendo $\rho=\min\left\{r,\frac{\lambda_{\bar{x}}}{2L}\right\}>0$ tenemos

$$\nabla^2 f(x)\in\mathbb{S}^n_{++}(\mathbb{R})\text{ con }\lambda_x\ge\frac{\lambda_{\bar{x}}}{2}>0,\quad x\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho).$$

De ese modo, para todo $x \in \mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$, existen matrices P_x y D_x tales que $\nabla^2 f(x) = P_x D_x P_x^\top$ con $P_x^{-1} = P_x^\top$, de modo que $\nabla^2 f(x)^{-1} = P_x D_x^{-1} P_x^\top$ y

$$\|\nabla^2 f(x)^{-1}\| = \sqrt{\lambda_{\text{máx}}(P_x D_x^{-2} P_x^\top)} \leq \frac{1}{\lambda_x} \leq \frac{2}{\lambda_{\bar{x}}}.$$

Supongamos que $x_k \in B_{\mathbb{R}^n}(\bar{x},\rho)$ para algún $k \in \mathbb{N}$. Para simplificar la notación, notemos $g_k = \nabla f(x_k)$ y $H_k = \nabla^2 f(x_k)$. De (4.15) se deduce que si $x_k = \bar{x}$ entonces $x_{k+1} = \bar{x}$, por lo que suponemos que $x_k \neq \bar{x}$. Como $\bar{x}$ es mínimo de f, usando el Teorema de Fermat, la propiedad de Lipschitz continuidad de $\nabla^2 f$ y la relación

$$g_k = \nabla f(x_k) - \nabla f(\bar{x}) = \int_0^1 \nabla^2 f(\bar{x} + t(x_k - \bar{x}))(x_k - \bar{x}) dt,$$

tenemos que

$$\begin{aligned}
|x_{k+1} - \bar{x}| &= |x_k - \bar{x} - H_k^{-1} g_k| \\
&= |H_k^{-1} (H_k(x_k - \bar{x}) - g_k)| \\
&= \left| H_k^{-1} \left(\int_0^1 [H_k - \nabla^2 f(\bar{x} + t(x_k - \bar{x}))](x_k - \bar{x}) dt \right) \right| \\
&\leq \frac{2}{\lambda_{\bar{x}}} |x_k - \bar{x}| \int_0^1 |H_k - \nabla^2 f(\bar{x} + t(x_k - \bar{x}))| dt \\
&\leq \frac{2L}{\lambda_{\bar{x}}} |x_k - \bar{x}|^2 \int_0^1 (1-t) dt \\
&= \frac{L}{\lambda_{\bar{x}}} |x_k - \bar{x}|^2 \leq \frac{1}{2} |x_k - \bar{x}|,
\end{aligned}$$

En particular, se tiene que $x_{k+1} \in \mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$. Además, usando inducción vemos que la sucesión $\{x_k\}$ está contenida en $\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$ si $x_0 \in \mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$ y

$$|x_{k+1} - \bar{x}| \leq \frac{1}{2^{k+1}} |x_0 - \bar{x}|, \quad \forall k \in \mathbb{N}.$$

De aquí se concluye que $x_k \to \bar{x}$, y que también tenemos

$$\frac{|x_{k+1} - \bar{x}|}{|x_k - \bar{x}|} \leq \frac{L}{\lambda_{\bar{x}}} |x_k - \bar{x}| \quad \text{y} \quad \frac{|x_{k+1} - \bar{x}|}{|x_k - \bar{x}|^2} \leq \frac{L}{\lambda_{\bar{x}}},$$

lo que finaliza la demostración. □

Observación 4.2. *El método también funciona si se asume que $\nabla^2 f$ es uniformemente continua en una vecindad acotada de $\bar{x}$, es decir, para algún $r > 0$ tenemos que*

$$\forall \varepsilon > 0,\ \exists \rho > 0,\ \forall x, y \in \mathbb{B}_{\mathbb{R}^n}(\bar{x}, r): \quad |x - y| \leq \rho \quad \Rightarrow \quad \|\nabla^2 f(x) - \nabla^2 f(y))\| \leq \varepsilon.$$

En ese caso la convergencia es superlineal:

$$\lim_{k \to \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = 0.$$

Por otra parte, el método no necesita que la función sea convexa en todo su dominio, ya que la demostración es local. Sin embargo, el método si necesita *que el mínimo sea único y que la función sea estrictamente convexa en una vecindad del mínimo.*

En el Ejemplo 4.5.2, el punto inicial no tiene relevancia para la convergencia. Sin embargo, en general la convergencia es garantizada sólo si se parte *suficientemente cerca* de la solución. El siguiente ejemplo ilustra un caso en que el método puede diverger si se parte lejos de la solución.

Ejemplo 4.5.3. *Sea* $f : \mathbb{R} \to \mathbb{R}$ *definida por*

$$f(x) = x\arctan(x) - \frac{1}{2}\ln(1+x^2), \quad \forall x \in \mathbb{R}.$$

Notemos que $f'(x) = \arctan(x)$ *que es estrictamente creciente, por lo que* f *es estrictamente convexa y el único mínimo se alcanza en* $\bar{x} = 0$*. Además, tenemos que* $f''(x) = \frac{1}{1+x^2}$ *por lo que, dado* $x_0 \in \mathbb{R}$*, la iteración del método de Newton-Raphson se escribe*

$$x_{k+1} = x_k - (1+x_k^2)\arctan(x_k).$$

En particular, si $x_0 = 10$ *tenemos la siguiente tabla con los términos de las iteraciones:*

k	x_k
0	10
1	-139
2	29892
3	-1403526593

Cuadro 4.2: Iteraciones del método Newton-Raphson

4.6. Ejercicios

1. FUNCIÓN CONVEXA DEFINIDA POR UNA INTEGRAL

   Consideremos el polinomio trigonométrico $T : \mathbb{R}^n \to [0,2\pi] \to \mathbb{R}$ definido por

   $$T(x,w) = x_1 + x_2\cos(w) + x_3\cos(2w) + \ldots + x_n\cos((n-1)w).$$

   Muestre que la función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ definida por

   $$f(x) = \begin{cases} -\displaystyle\int_0^{2\pi} \log(T(x,w))dw & \text{si } T(x,w) > 0, \forall w \in [0,2\pi], \\ +\infty & \text{si no} \end{cases}$$

   es una función convexa.

2. FUNCIÓN CONVEXA VECTORIAL-MATRICIAL

   Se define la función $f : \mathbb{R}^n \times \mathbb{S}^n(\mathbb{R}) \to \mathbb{R}$ como sigue

   $$f(x,A) = \begin{cases} x^\top A^{-1}x & \text{si } x \in \mathbb{R}^n,\ A \in \mathbb{S}^n_{++}(\mathbb{R}), \\ +\infty & \text{si no} \end{cases}$$

   a) Muestre que $\operatorname{dom}(f)$ es un abierto de $\mathbb{R}^n \times \mathbb{S}^n(\mathbb{R})$ y que f es Gâteaux diferenciable con

   $$\nabla f(x,A)(d,D) = 2x^\top A^{-1}d - x^\top A^{-1}DA^{-1}x, \quad \forall x \in \mathbb{R}^n,\ \forall D \in \mathbb{S}^n(\mathbb{R}).$$

   Aquí suponemos que $\mathbb{S}^n(\mathbb{R})$ tiene la estructura de espacio de Hilbert con producto interno usual: $\langle A,B\rangle = \operatorname{tr}(AB)$ para todo $A,B \in \mathbb{S}^n(\mathbb{R})$

   b) Deducir que f es una función convexa demostrando que f es subdiferenciable, es decir,

   $$f(x,A) + \nabla f(x,A)(y-x, B-A) \le f(y,B), \quad \forall (x,A),(y,B) \in \operatorname{dom}(f).$$

   Indicación: Calcular $(A^{-1}x - B^{-1}y)^\top B(A^{-1}x - B^{-1}y)$.

3. CONDICIONES DE OPTIMALIDAD PARA FUNCIONES NO DIFERENCIABLES

   Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert. Considere $g,h : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ dos funciones convexas y propias con $\operatorname{dom}(g) \cap \operatorname{dom}(h) \neq \emptyset$. Suponga que g es Gâteaux diferenciable en $\operatorname{dom}(g)$, es último siendo un abierto de $\mathbf{X}$. Definamos la función $f : \mathbf{X} \to \mathbb{R}$ por

   $$f(x) = g(x) + h(x), \qquad \forall x \in \mathbf{X}.$$

   a) Pruebe que $\bar{x} \in \arg\min_{\mathbf{X}}(f)$ si y sólo si

   $$\langle \nabla g(\bar{x}), x - \bar{x}\rangle + h(x) - h(\bar{x}) \ge 0, \quad \forall x \in \mathbf{X}.$$

   b) Muestre además que si $x \mapsto \nabla g(x)$ es secuencialmente fuerte-$\sigma(\mathbf{X},\mathbf{X})$ continuo en $\operatorname{dom}(g)$, es decir, para cualquier $\{x_k\} \subseteq \operatorname{dom}(g)$, si $x_k \to x \in \operatorname{dom}(g)$ se tiene que

   $$\nabla g(x_k) \xrightarrow[k\to\infty]{} \nabla g(x),$$

   entonces $\bar{x} \in \arg\min_{\mathbf{X}}(f)$ si y sólo si

   $$\langle \nabla g(x), x - \bar{x}\rangle + h(x) - h(\bar{x}) \ge 0, \quad \forall x \in \operatorname{dom}(g).$$

4. DESIGUALDAD DE KANTOROVICH

   Sea $A \in \mathbb{S}^n_{++}(\mathbb{R})$ con valores propios $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n 0$. El objetivo de esta pregunta es demostrar la desigualdad

$$|x|^4 \leq \left(x^\top Ax\right)\left(x^\top A^{-1}x\right) \leq \frac{1}{4}\left[\sqrt{\frac{\lambda_n}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_n}}\right]^2 |x|^4, \qquad \forall x \in \mathbb{R}^n.$$

   Para ello se aconseja

   a) Si $A = P^\top DP$ es una diagonalización de A, demostrar que para obtener la desigualdad basta probar

$$1 \leq \left(y^\top Dy\right)\left(y^\top D^{-1}y\right) \leq \frac{1}{4}\left[\sqrt{\frac{\lambda_n}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_n}}\right]^2, \quad \forall y \in \mathbb{R}^n \text{ con } |y| = 1.$$

   b) Defina $\bar{\lambda} = \sum_{i=1}^n y_i^2 \lambda_i$ y pruebe que

$$\frac{1}{\bar{\lambda}} \leq \sum_{i=1}^n \frac{y_i^2}{\lambda_i} \leq \frac{\lambda_1 + \lambda_n - \bar{\lambda}}{\lambda_1 \lambda_n},$$

   y a partir de esto obtenga el resultado buscado.

5. MÉTODO DE NEWTON-RAPHSON Y PROBLEMAS CUADRÁTICOS

   Considere la función $f\colon \mathbb{R}^n \to \mathbb{R}$ definida por

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

   donde $A \in \mathbb{S}^n_{++}(\mathbb{R})$, $b \in \mathbb{R}^n$ y $c \in \mathbb{R}$. Pruebe que para cualquier $x_0 \in \mathbb{R}^n$, el método de Newton-Raphson aplicado a la función f converge en solo una iteración.

6. FORMA ALTERNATIVA DEL MÉTODO GRADIENTE CONJUGADO

   Dados $A \in \mathbb{S}^n_{++}(\mathbb{R})$, $b \in \mathbb{R}^n$ y $c \in \mathbb{R}$, considere la función cuadrática $f\colon \mathbb{R}^n \to \mathbb{R}$ definida por

$$f(x) = \frac{1}{2}x^\top Ax - b^\top x + c, \quad \forall x \in \mathbb{R}^n.$$

   Dado $x_0 \in \mathbb{R}^n$ y $g_0 = \nabla f(x_0) = Ax - b$, considere para todo $k \in \{1, \ldots, n\}$, el punto x_{k+1} que se encuentra resolviendo el problema

$$\text{Minimizar } f(x) \text{ sobre todos los } x \in U_k,$$

   donde $U_k = \{x_k\} + V_k$ y V_k es el espacio vectorial generado por $g_0, \ldots, g_k$. Demuestre que el método equivale al Método del Gradiente Conjugado, es decir, que cada x_k es la $k-$ésima iteración del Método del Gradiente Conjugado que parte desde x_0.

CAPÍTULO 5

Optimización convexa no diferenciable

Abstract. En este capítulo estudiaremos funciones convexas no diferenciables y veremos que la Regla de Fermat tiene un análogo si reemplazamos el diferencial por una noción generalizada de este, el cual llamaremos subdiferencial. Usaremos esta nueva herramienta para obtener condiciones de optimalidad para problemas con restricciones y estudiaremos algunos métodos para resolver esta clase de problemas.

Recordemos que una forma de estudiar problemas de optimización con restricciones es incluir en la definición de la función objetivo la restricción via una penalización fuerte. Dicho de otra forma, resolver

(P) Minimizar $f(x)$ sobre todos los $x \in \mathbf{X}$ que satisfacen la restricción $x \in \mathbf{S}$

donde $f : \mathbf{X} \to \mathbb{R}$ es una función dada y $\mathbf{S} \subseteq \mathbf{X}$ un conjunto dado, es equivalente a resolver

($\mathrm{P}_{\mathbf{S}}$) Minimizar $f_{\mathbf{S}}(x) := f(x) + \delta_{\mathbf{S}}(x)$ sobre todos los $x \in \mathbf{X}$.

Notemos que en caso que (P) sea un problema convexo, tendremos que $f_{\mathbf{S}}$ será también una función convexa. Es importante destacar que no importa la regularidad que impongamos sobre f, la función $f_{\mathbf{S}}$ no será jamás diferenciable en la frontera de $\mathbf{S}$ (salvo en el caso trivial $\mathbf{S} = \mathbf{X}$), lo cual en principio no nos permitiría aplicar los resultados vistos en el capítulo anterior a funciones similares a $f_{\mathbf{S}}$. Afortunadamente, para el caso de optimización convexa, la diferenciabilidad es una herramienta útil pero no fundamental, pues mucho resultados pueden ser extendidos al caso no diferenciable introduciendo un objeto matemático llamado *subdiferencial*.

En este capítulo, y sólo con el propósito de simplificar la exposición, trabajaremos básicamente con funciones $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ convexas definidas sobre un espacio de Hilbert $(\mathbf{X}, \langle \cdot, \cdot \rangle)$. La condición impuesta anteriormente, que $\mathrm{dom}(f)$ sea un abierto de $\mathbf{X}$, no será necesaria a partir de ahora.

5.1. Subdiferencial

El concepto de subdiferencial viene a generalizar la idea del diferencial de una función. La definición calza bien para funciones convexas, sin embargo hay que notar que ésta no requiere en absoluto de la convexidad de la función en cuestión.

Definición 5.1. *Supongamos que $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ es un espacio de Hilbert y sea $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función dada. Un subgradiente de f en $x \in \mathbf{X}$ es un vector $x^* \in \mathbf{X}$ que satisface*

$$f(x) + \langle x^*, y - x \rangle \leq f(y), \quad \forall y \in \mathbf{X}.$$

La colección de todos los subgradientes de f en x, denotada $\partial f(x)$, es el subdiferencial de f en x.

La idea esencial del subdiferencial es agrupar todas las posibles pendientes que pueden tener las funciones afines continuas que minoran a la función convexa en cuestión.

Observación 5.1. *Notemos que $\partial f(x)$ es un conjunto convexo, posiblemente vacío, y cerrado para la topología débil en $\mathbf{X}$ (y por lo tanto cerrado para la topología fuerte de $\mathbf{X}$), cualquiera sea $x \in \mathbf{X}$. Además, es claro que si f es propia, $\partial f(x) = \emptyset$ cada vez que $f(x) = +\infty$.*

Ejemplo 5.1.1. *Veamos algunos ejemplos:*

- *Sea $f(x) = |x|$ para cada $x \in \mathbb{R}$, entonces $\partial f(0) = [-1, 1]$.*
- *Sea $f(x) = -\sqrt{x} + \delta_{[0,+\infty)}(x)$ para cada $x \in \mathbb{R}$, entonces $\partial f(0) = \emptyset$.*

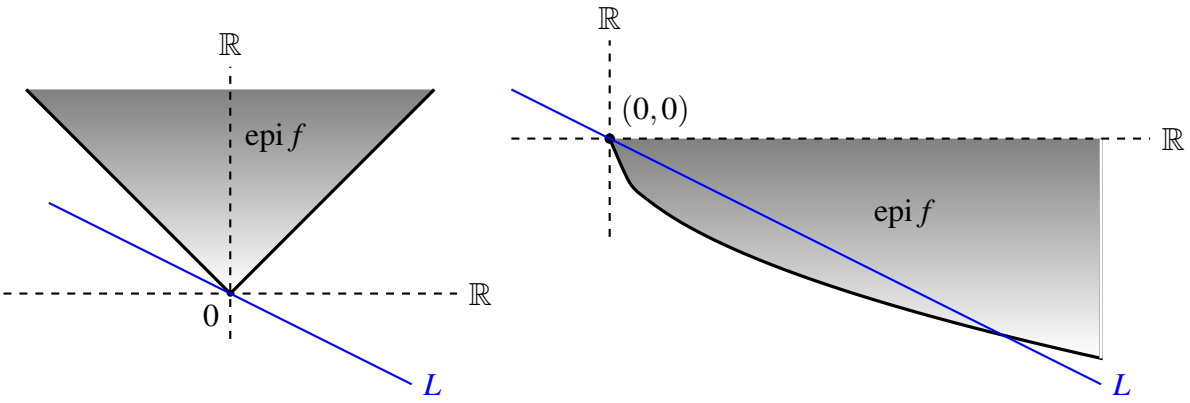


Figura 5.1: Epígrafo de las funciones $f(x) = |x|$ y $f(x) = -\sqrt{x} + \delta_{[0,+\infty)}(x)$.

Como muestra uno de los ejemplos anteriores, el subdiferencial de una función convexa puede ser vacío, incluso si la función es finita en el punto. Un criterio, relativamente simple, para evitar esto es que la función sea continua. Esto es una consecuencia del Teorema de Hahn-Banach.

Proposición 5.1. *Sean $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función convexa. Suponga que f es continua en $x \in dom(f)$, entonces $\partial f(x) \neq \emptyset$*

Demostración. Dado que f es continua en x, podemos encontrar $r > 0$ tal que $f(x + rd) \leq f(x) + 1$ para cada $d \in \mathbb{B}_{\mathbf{X}}$. De donde se tiene que $\text{int}(\text{epi}(f)) \neq \emptyset$ y además $(x, f(x)) \notin \text{int}(\text{epi}(f))$. Luego por el Teorema de Hahn-Banach (Lema 3.1), existe $(x^*, \alpha) \in \mathbf{X} \times \mathbb{R} \setminus \{0\}$ tal que

$$\langle x^*, x \rangle + \alpha f(x) \leq \langle x^*, y \rangle + \alpha\lambda, \quad \forall (y, \lambda) \in \text{epi}(f)$$

De aquí se concluye que $\alpha \geq 0$ pues $(x, \lambda) \in \text{epi}(f)$ para cualquier $\lambda \geq f(x)$. Además, como $x + rd \in \text{dom}(f)$ para cada $d \in \mathbb{B}_{\mathbf{X}}$, tenemos que si $\alpha = 0$ entonces

$$\langle x^*, x \rangle \leq \langle x^*, x + rd \rangle, \quad \forall d \in \mathbb{B}_{\mathbf{X}}.$$

Esto a su vez implica que $\|x^*\|_* = 0$ y por lo tanto $(x^*, \alpha) = 0$, llevándonos a una contradicción. Por lo tanto $\alpha > 0$, y sin perdida de generalidad podemos asumir que $\alpha = 1$, multiplicando x^* por $\frac{1}{\alpha}$ si es necesario. Entonces, tenemos que

$$\langle x^*, x - y \rangle + f(x) \leq \lambda, \quad \forall (y, \lambda) \in \text{epi}(f).$$

Tomando $\lambda = f(y)$, vemos que $x^* \in \partial f(x)$, y la proposición ha sido demostrada. □

5.1.1. Cono Normal

Un ejemplo interesante a estudiar es el subdiferencial de la función indicatriz $f = \delta_{\mathbf{S}}$ donde $\mathbf{S} \subseteq \mathbf{X}$ es un conjunto dado. El conjunto $\partial\delta_{\mathbf{S}}(x)$ se conoce como el *cono normal* a $\mathbf{S}$ en $x \in \mathbf{X}$ y viene dado por

$$N_{\mathbf{S}}(x) := \partial\delta_{\mathbf{S}}(x) = \{x^* \in \mathbf{X} \mid \langle x^*, y - x\rangle \leq 0,\ \forall y \in \mathbf{S}\}.$$

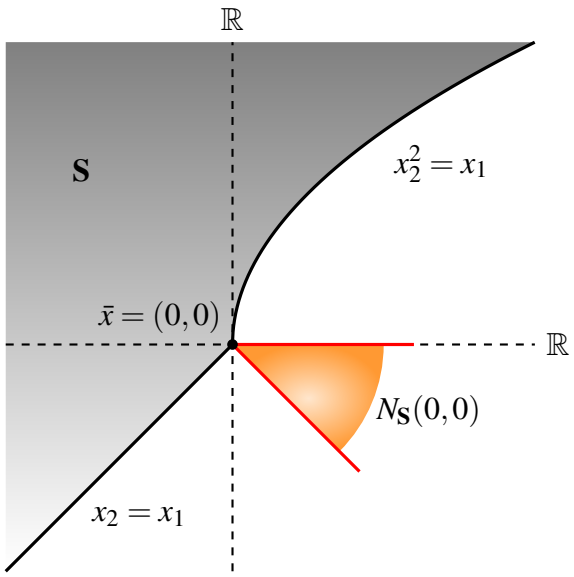


Figura 5.2: Ejemplo cono normal a un conjunto en $\mathbb{R}^2$.

El cono normal jugará un rol importante cuando escribamos condiciones de optimalidad. En particular, será de importancia conocer la estructura del cono normal a un conjunto de nivel, es decir, $\mathbf{S} = \Gamma_\gamma(f)$ para cierto $\gamma \in \mathbb{R}$ y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$. A continuación daremos una respuesta parcial a la estructura del cono normal en este caso. Para demostrar el converso de la siguiente proposición necesitamos algunas herramientas que aún no tenemos, por lo cual posponemos esa parte de la demostración para más adelante; ver Proposición 5.6.

Proposición 5.2. *Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función propia tal que $\Gamma_\gamma(f) \neq \emptyset$ para cierto $\gamma \in \mathbb{R}$. Luego, tenemos que*

$$\forall x \in \Gamma_\gamma(f),\ \forall \mu \geq 0,\ \forall x^* \in \partial f(x) \text{ tales que } \mu(f(x) - \gamma) = 0 \text{ se tiene que } \mu x^* \in N_{\Gamma_\gamma(f)}(x).$$

Demostración. Notemos que para cada $x^* \in \partial f(x)$ y $\mu \geq 0$ tenemos que

$$\langle \mu x^*, y - x\rangle + \mu f(x) \leq \mu f(y), \quad \forall y \in \mathbf{X}.$$

Por lo tanto, si $y \in \Gamma_\gamma(f)$, obtenemos la desigualdad

$$\langle \mu x^*, y - x\rangle \leq \mu(f(y) - f(x)) \leq \mu(\gamma - f(x)).$$

De aquí se concluye que si $\mu(f(x) - \gamma) = 0$ se tendrá también que $\mu x^* \in N_{\Gamma_\gamma(f)}(x)$. □

5.1.2. Relación con diferenciabilidad

La relación que existe entre el subdiferencial y el diferencial de una función puede ser estudiada a través de la derivada direccional. Recordemos que la esta derivada está dada por

$$f'(x;d) := \lim_{t\to 0^+} \frac{f(x+td)-f(x)}{t}, \quad \forall d \in \mathbf{X}.$$

Notemos que en general $-\infty \leq f'(x;d) \leq +\infty$. De hecho los valores $\pm\infty$ pueden ser alcanzados, como lo muestra el siguiente ejemplo.

Ejemplo 5.1.2. *Consideremos la función $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ definida por*

$$f(x) = \begin{cases} -\sqrt{1-x^2} & si\ |x| \leq 1 \\ +\infty & si\ no \end{cases}.$$

Luego se tiene que $f'(-1,d) = -\infty$ y $f'(1,d) = +\infty$ para cada $d > 0$. En efecto

$$f'(-1,d) = \lim_{t\to 0^+} \frac{-\sqrt{1-(-1+td)^2}-0}{t} = \lim_{t\to 0^+} -\sqrt{\frac{2td-t^2d^2}{t^2}} = \lim_{t\to 0^+} -\sqrt{\frac{2d}{t}-d^2} = -\infty$$

y dado que $1+td > 1$ si $t,d > 0$ entonces se tiene que

$$f'(1,d) = \lim_{t\to 0^+} \frac{f(1+td)-0}{t} = +\infty$$

La derivada direccional es importante en Análisis Convexo pues permite obtener una representación del subdiferencial de una función convexa, como veremos a continuación.

Proposición 5.3. *Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert, $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función convexa y $x \in dom(f)$. Entonces*

$$f'(x;d) = \inf_{t>0} \frac{f(x+td)-f(x)}{t}, \qquad \forall d \in \mathbf{X}.$$

Además, $d \mapsto f'(x;d)$ es sublineal y

$$\partial f(x) = \{x^* \in \mathbf{X} \mid \ \langle x^*, d\rangle \leq f'(x;d),\ \forall d \in \mathbf{X}\}, \qquad \forall x \in \mathbf{X}.$$

Demostración. Sea $x \in \mathrm{dom}(f)$. Dividamos la demostración en partes:

1. Consideremos $d \in \mathbf{X}$ y la función $g(t) = \frac{f(x+td)-f(x)}{t}$ definida para todo $t \in (0,+\infty)$. Veamos que para $0 < t \leq s$ se tiene que $g(t) \leq g(s)$. Notemos que

$$x+td = \frac{t}{s}(x+sd) + \left(1-\frac{t}{s}\right)x.$$

Dado que f es convexa, sigue que

$$f(x+td) \leq \frac{t}{s}f(x+sd) + \left(1-\frac{t}{s}\right)f(x).$$

de donde se concluye que

$$g(t) = \frac{f(x+td)-f(x)}{t} \leq \frac{f(x+sd)-f(x)}{s} = g(s).$$

De este modo, como $t \mapsto g(t)$ es creciente en $(0,+\infty)$, se tiene

$$f'(x,d) = \lim_{t\to 0^+} \frac{f(x+td)-f(x)}{t} = \lim_{t\to 0^+} g(t) = \inf_{t>0} g(t) = \inf_{t>0} \frac{f(x+td)-f(x)}{t}.$$

2. Veamos ahora que $d \mapsto f'(x;d)$ es sublineal, es decir,

$$f'(x;d_1+d_2) \leq f'(x;d_1)+f'(x;d_2), \qquad \forall d_1,d_2 \in \mathbf{X}.$$

Notemos que si $f'(x;d_1+d_2) = -\infty$, $f'(x;d_1) = +\infty$ o $f'(x;d_2) = +\infty$, entonces el resultado es trivial; recordando las convenciones que hemos aceptado. Luego, asumamos que

$$f'(x;d_1+d_2) > -\infty, \quad f'(x;d_1) < +\infty \quad \text{y} \quad f'(x;d_2) < +\infty.$$

Dado que f es convexa, la parte anterior implica que

$$f'(x;d_1+d_2) \leq \frac{f(x+t(d_1+d_2))-f(x)}{t} \leq \frac{\frac{1}{2}f(x+2td_1)+\frac{1}{2}f(x+2td_2)-f(x)}{t}, \quad \forall t>0,$$

y por lo tanto, haciendo un cambio de variable ($2t$ por t) tenemos que

$$f'(x;d_1+d_2) \leq \frac{f(x+td_1)-f(x)}{t} + \frac{f(x+td_2)-f(x)}{t}, \quad \forall t>0.$$

Luego como $f'(x;d_1) < +\infty$, podemos encontrar un $s_1 > 0$ tal que $\frac{f(x+s_1d_1)-f(x)}{s_1} < +\infty$. Esto a su vez, junto con la monotonía del cuociente, implica que

$$f'(x;d_1+d_2) \leq \frac{f(x+s_1d_1)-f(x)}{s_1} + \frac{f(x+td_2)-f(x)}{t}, \quad \forall t \in (0,s_1],$$

de donde obtenemos, al tomar ínfimo sobre t, que $f'(x;d_2) > -\infty$. Tomemos ahora $\varepsilon > 0$, por definición de ínfimo podemos encontrar $s_2 > 0$ para el cual $\frac{f(x+s_2d_2)-f(x)}{s_2} \leq f'(x;d_2)+\varepsilon$. Sigue que por la monotonía del cuociente tenemos

$$f'(x;d_1+d_2) \leq \frac{f(x+td_1)-f(x)}{t} + f'(x;d_2)+\varepsilon, \quad \forall t \in (0,s_2].$$

Finalmente, tomando ínfimo sobre t en la desigualdad anterior llegamos a la conclusión, ya que $\varepsilon > 0$ es un número positivo arbitrario.

3. Tomemos ahora $x^* \in \partial f(x)$ y $d \in \mathbf{X}$ arbitrario. La definición del subdiferencial nos lleva a

$$\langle x^*, d\rangle \leq \frac{f(x+td)-f(x)}{t}, \quad \forall t \in (0,+\infty).$$

De esta desigualdad se concluye fácilmente que $\langle x^*,d\rangle \leq f'(x,d)$. Por otra parte, tomemos $x^* \in \mathbf{X}$ tal que $\langle x^*,d\rangle \leq f'(x,d)$ para todo $d \in \mathbf{X}$. Usando la primera parte con $t = 1$ tenemos

$$\langle x^*,d\rangle \leq f'(x,d) \leq f(x+d)-f(x).$$

Finalmente, tomando $d = y-x$ con $y \in \mathrm{dom}(f)$ arbitrario se concluye que $x^* \in \partial f(x)$. Esto entrega la caracterización del subdiferencial y termina la demostración.

□

Ahora veremos que la relación entre el subdiferencial y la derivada direccional de una función convexa es unívoca, en el sentido que la derivada direccional puede ser calculada a partir del subdiferencial. El siguiente resultado mostrará en particular que una función convexa es diferenciable si y sólo si el subdiferencial tiene un único elemento. Cabe destacar que el resultado que veremos a continuación es una consecuencia de la versión analítica del Teorema de Hahn-Banach, la cual es equivalente a la versión geométrica de este teorema (Lema 3.1); a partir de uno se puede demostrar el otro.

Recuerdo : Teorema analítico de Hahn-Banach

La versión analítica del Teorema de Hahn-Banach dice que un funcional lineal continuo definido solo en un subespacio de $\mathbf{X}$ que satisface una cota apropiada, puede ser extendido a todo el espacio, satisfaciendo la misma cota. Por esta razón, mucha veces el teorema se conoce como el Teorema de extensión de Hahn-Banach.

Lema 5.1 (Teorema Hahn-Banach Analítico)**.** *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert y* $g : \mathbf{X} \to \mathbb{R}$ *una función sublineal y positivamente homogénea, es decir,*

$$g(x+y) \leq g(x)+g(y) \quad y \quad g(\alpha x) = \alpha g(x), \quad \forall x,y \in \mathbf{X},\ \forall \alpha > 0.$$

Sea $\mathbf{X}_0$ *un subespacio vectorial de* $\mathbf{X}$ *y* $\ell_0 : \mathbf{X}_0 \to \mathbb{R}$ *un funcional lineal tal que*

$$\ell_0(x) \leq g(x), \quad \forall x \in \mathbf{X}_0.$$

Entonces, existe $\ell : \mathbf{X} \to \mathbb{R}$ *lineal tal que*

$$\ell(x) = \ell_0(x), \quad \forall x \in \mathbf{X}_0 \quad y \quad \ell(x) \leq g(x), \quad \forall x \in \mathbf{X}.$$

Proposición 5.4. *Sean* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función convexa. Suponga que* f *es continua en* $x \in dom(f)$*. Entonces*

$$f'(x;d) = \text{máx}\left\{\langle x^*, d\rangle \mid x^* \in \partial f(x)\right\}, \quad \forall d \in \mathbf{X}$$

Además, se tiene que f *es Gâteaux diferenciable en* x *si y sólo si* $\partial f(x) = \{x^*\}$.

Demostración. Notemos primero que gracias a la Proposición 5.3, la desigualdad " $\geq$ " siempre es cierta; independiente de la continuidad de f. Luego bastará probar que existe $x^* \in \partial f(x)$ tal que

$$f'(x;d) = \langle x^*, d\rangle, \quad \forall d \in \mathbf{X}.$$

Notemos además que gracias a Proposición 5.1, tenemos que $\partial f(x) \neq \emptyset$. Fijemos $d \in \mathbf{X} \setminus \{0\}$.

Consideremos el espacio vectorial $\mathbf{X}_0 = \{\alpha d \mid \alpha \in \mathbb{R}\}$ y la función lineal $\ell_0 : \mathbf{X}_0 \to \mathbb{R}$ definida por

$$\ell_0(\alpha d) = \alpha f'(x;d), \quad \forall \alpha \in \mathbb{R}.$$

Notemos que si $\alpha > 0$, entonces

$$\ell_0(\alpha d) = \alpha \inf_{t>0} \frac{f(x+td)-f(x)}{t} = \inf_{t>0} \frac{f(x+\frac{t}{\alpha}\alpha d)-f(x)}{\frac{t}{\alpha}} = f'(x;\alpha d).$$

De aquí no es difícil ver que $v \mapsto g(v) := f'(x;v)$ es positivamente homogénea. Además, si $\alpha < 0$, entonces

$$\ell_0(\alpha d) = \alpha f'(x;d) = -f'(x;-\alpha d) \leq f'(x;\alpha d),$$

donde la última desigualdad viene del hecho que $d \mapsto f'(x;d)$ es sublineal y $f'(x;0) = 0$. Luego por el Teorema de extensión de Hahn-Banach (Lema 5.1), existe un funcional lineal $\ell : \mathbf{X} \to \mathbb{R}$ tal que

$$\ell(d) = f'(x;d) \quad \text{y} \quad \ell(v) \leq f'(x;v), \quad \forall v \in \mathbf{X}.$$

Tomando $v = y - x$ para cualquier $y \in \text{dom}(f)$ y usando Proposición 5.3, vemos que

$$\ell(y-x) \leq f'(x;y-x) \leq f(y) - f(x).$$

Luego para concluir basta ver que ℓ es continuo, y que por lo tanto existe $x^* \in \mathbf{X}$ tal que $\ell = \langle x^*, \cdot \rangle$. Dado que f es continuo en x, se tiene que para todo $\varepsilon > 0$, existe $r > 0$ tal que $|f(x) - f(y)| \leq \varepsilon$ para todo $y \in \mathbb{B}_{\mathbf{X}}(x,r)$. Luego tenemos, por la desigualdad del subdiferencial que

$$\ell(y-x) \leq |f(x) - f(y)| \leq \varepsilon, \quad \forall y \in \mathbb{B}_{\mathbf{X}}(x,r).$$

Evaluando en $2x - y$ en vez de en y, se obtiene la desigualdad con el valor absoluto. Esto implica que ℓ es continuo en x, pero al ser lineal, debe ser continuo en todo punto de $\mathbf{X}$ y por lo tanto existe $x^* \in \mathbf{X}$ tal que $\ell = \langle x^*, \cdot \rangle$. Esto concluye la demostración del resultado.

□

5.1.3. Reglas de cálculo

Llegamos al punto en que podemos presentar un análogo de la Regla de Fermat para el caso no diferenciable, simplemente reemplazando el diferencial por el subdiferencial. Notemos que en el siguiente teorema la convexidad de la función objetivo no es necesaria.

Teorema 5.1 (Regla de Fermat II)**.** *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función propia. Entonces*

$$\bar{x} \in \arg\min_{\mathbf{X}}(f) \quad \Longleftrightarrow \quad 0 \in \partial f(\bar{x})$$

Demostración. Directo de la definición del subdiferencial. □

Observación 5.2. *En la práctica, para encontrar un mínimo se necesita probar que $0 \in \partial f(\bar{x})$ para algún $\bar{x} \in \mathbf{X}$. Es en esta parte donde la convexidad de la función juega un rol esencial.*

Regla de la suma

Como mencionamos anteriormente, en muchas ocaciones estamos interesados en encontrar mínimos de una función que se puede escribir como la suma de dos funciones convexas, con al menos una de ella no diferenciable; por ejemplo funciones del tipo $f_{\mathbf{S}} := f + \delta_{\mathbf{S}}$. Por esta razón es importante proveer una regla para calcular el subdiferencial de la suma de funciones convexas.

Teorema 5.2 (Moreau-Rockafellar)**.** *Sean $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y $f_1, f_2 : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ funciones propias convexas y s.c.i.. Supongamos que f_1 es continua en $x_0 \in dom(f_1) \cap dom(f_2)$. entonces*

$$\partial f_1(x) + \partial f_2(x) = \partial(f_1 + f_2)(x), \quad \forall x \in \mathbf{X}.$$

Demostración. Comencemos la demostración probando la inclusión ($\subseteq$) que resulta de la definición. Efectivamente, tenemos que si $x_1^* \in \partial f_1(x)$ y $x_2^* \in \partial f_2(x)$ entonces

$$f_1(x) + \langle x_1^*, y - x\rangle \leq f_1(y), \quad \forall y \in \mathbf{X},$$
$$f_2(x) + \langle x_2^*, y - x\rangle \leq f_2(y), \quad \forall y \in \mathbf{X}.$$

Luego sumando ambas desigualdades se obtiene que $x_1 + x_2^* \in^* \in \partial(f_1 + f_2)(x)$

Probemos ahora la otra inclusión ($\supseteq$), la cual requiere un poco más de desarrollo. Sean $x \in \mathbf{X}$ y $x^* \in \mathbf{X}$ tales que $x^* \in \partial(f_1 + f_2)(x)$. Tenemos por definición que

$$\langle x^*, y - x\rangle + f_1(x) + f_2(x) \leq f_1(y) + f_2(y), \quad \forall y \in \mathbf{X}. \tag{5.1}$$

Introduzcamos los siguientes conjuntos convexos

$$A := \{(y,\lambda) \in \mathbf{X} \times \mathbb{R} \mid f_1(y) - \langle x^*, y - x\rangle \leq \lambda\} \quad \text{y} \quad B := \{(y,\lambda) \in \mathbf{X} \times \mathbb{R} \mid f_1(x) + f_2(x) - f_2(y) \geq \lambda\}.$$

Notemos que $(y,\lambda) \in A \cap B$ es equivalente a pedir que

$$f_1(y) + f_2(y) \leq \langle x^*, y - x\rangle + f_1(x) + f_2(x),$$

la cual es en realidad una igualdad debido a (5.1).

Por otro lado vemos que $A = \operatorname{epi}(g)$ con $g = f_1 - \langle x^*, \cdot - x\rangle$, la cual es una función propia convexa y s.c.i., que además es continua en x_0. Luego, $\emptyset \neq \operatorname{int}(A) \subseteq \{(y,\lambda) \in \mathbf{X} \times \mathbb{R} \mid g(y) < \lambda\}$ y $\operatorname{int}(A)$ es convexo. Más aún, $\operatorname{int}(A) \cap B = \emptyset$, y por lo tanto podemos separar $\operatorname{int}(A)$ de B mediante un hiperplano cerrado gracias al Teorema de Hahn-Banach (Lema 3.1). En otras palabras, $\exists (y^*, r) \in \mathbf{X} \times \mathbb{R} \setminus \{0\}$ y $\alpha \in \mathbb{R}$ tales que

$$\langle y^*, y\rangle + r\lambda < \alpha, \ \forall (y,\lambda) \in \operatorname{int}(A) \quad \text{y} \quad \langle y^*, \tilde{y}\rangle + r\tilde{\lambda} \geq \alpha, \ \forall (\tilde{y},\tilde{\lambda}) \in B.$$

Notemos que $(x,\lambda) \in B$ si y sólo si $\lambda \leq f_1(x)$, y por lo tanto r no puede ser positivo. Además, como $(x_0, g(x_0) + \varepsilon) \in \operatorname{int}(A)$ para algún $\varepsilon > 0$ debemos necesariamente tener que $r < 0$. En efecto, si $r = 0$ y dado que $(x_0, f_1(x) + f_2(x) - f_2(x_0)) \in B$, entonces tendríamos

$$\langle y^*, x_0\rangle < \alpha \leq \langle y^*, x_0\rangle,$$

lo que no puede ser. Por lo tanto debemos tener que

$$\langle -x_2^*, y\rangle - \lambda < \langle -x_2^*, \tilde{y}\rangle - (f_1(x) + f_2(x) - f_2(\tilde{y})), \quad \forall (y,\lambda) \in \operatorname{int}(A), \ \tilde{y} \in \operatorname{dom}(f_2)$$

donde $x_2^* = \frac{1}{r} y^*$. Notemos que, para todo $y \in \operatorname{int}\operatorname{dom} f_1$ y $\lambda \in \mathbb{R}$ tales que $f_1(y) - \langle x^*, y\rangle < \lambda$, se tiene que $(y,\lambda) \in \operatorname{int}(A)$ y luego, haciendo $\lambda \to f_1(y) - \langle x^*, y - x\rangle$, obtenemos que

$$\langle -x_2^*, y\rangle - f_1(y) + \langle x^*, y - x\rangle \leq \langle -x_2^*, \tilde{y}\rangle - (f_1(x) + f_2(x) - f_2(\tilde{y})), \quad \forall y \in \operatorname{int}\operatorname{dom} f_1, \tilde{y} \in \operatorname{dom} f_2.$$

Tomando $\tilde{y} = x \in \operatorname{dom} f_2$, deducimos que

$$f_1(x) + \langle -x_2^* + x^*, y - x\rangle \leq f_1(y), \quad \forall y \in \operatorname{int}\operatorname{dom}(f_1),$$

lo que, en conjunto con la Proposición 3.1, implican que $-x_2^* + x^* \in \partial f_1(x)$ y análogamente, tomando $y = x$

$$f_2(x) + \langle x_2^*, y - x\rangle \leq f_2(y), \quad \forall y \in \operatorname{dom}(f_2),$$

de donde concluimos que $x_2^* \in \partial f_2(x)$. Definiendo $x_1^* = x^* - x_2^*$ se tiene que $x_1^* \in \partial f_1(x)$ y $x_1^* + x_2^* = x^*$ lo que termina la demostración. □

El siguiente es un contraejemplo que muestra que la igualdad no se tiene si la condición que alguna función sea continua en un punto común a ambos dominios no se satisface.

Ejemplo 5.1.3. *Supongamos que* $\mathbf{X} = \mathbb{R}^2$*, sean* $C_1 = \{(x,y) \in \mathbb{R}^2 \,|\, (x-1)^2 + y^2 \leq 1\}$, $C_2 = \{(x,y) \in \mathbb{R}^2 \,|\, (x+1)^2 + y^2 \leq 1\}$, $f_1 = \delta_{C_1}$ *y* $f_2 = \delta_{C_2}$*. Luego* $f_1 + f_2 = \delta_{(0,0)}$ *y* $\partial(f_1 + f_2)(0,0) = \mathbb{R}^2$*. Por otro lado, se tiene* $\partial f_1(0,0) =]-\infty, 0] \times \{0\}$, $\partial f_2(0,0) = [0, +\infty[\times \{0\}$*, de donde* $\partial f_1(0,0) + \partial f_1(0,0) = \mathbb{R} \times \{0\}$*. Notar que ninguna de las funciones es continua en* $\{(0,0)\} = dom\, f_1 \cap dom\, f_2$*.*

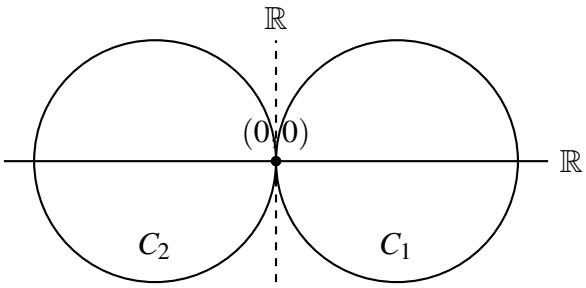


Figura 5.3: Contraejemplo regla de la suma.

Regla de la composición

Además de la Regla de la suma, el subdiferencial satisface un regla sobre la composición con operadores lineales, la cual es lo más cercano que podemos obtener a una regla de la cadena para subdiferenciales de funciones convexas. Esta regla será de particular utilidad para resolver problemas tales como el problema de *Compresión y recuperación de imágenes* (Sección 1.2), el cual tiene una estructura del tipo

$$\text{Minimizar } f(x) + g(Ax) \text{ sobre todos los } x \in \mathbf{X}$$

donde $A : \mathbf{X} \to \mathbf{Y}$ es un operador lineal continuo y $g : \mathbf{Y} \to \mathbb{R} \cup \{+\infty\}$ es una función convexa.

Recuerdo : Operador adjunto

Sea $A : \mathbf{X} \to \mathbf{Y}$ es un operador lineal continuo entre dos espacios de Hilbert $\mathbf{X}$ e $\mathbf{Y}$, se define el operador adjunto de A, denotado por $A^* : \mathbf{Y} \to \mathbf{X}$, como el operador lineal continuo que satisface

$$\langle y, Ax \rangle = \langle A^* y, x \rangle, \qquad \forall x \in \mathbf{X},\ y \in \mathbf{Y}.$$

En el caso que $\mathbf{X} = \mathbb{R}^n$ e $\mathbf{Y} = \mathbb{R}^m$, todo operador lineal continuo puede ser representado a través de una matriz. Luego, abusando de la notación se tiene que $A \in \mathbb{M}_{n\times m}(\mathbb{R})$ y el operador adjunto no es otra cosa que la matriz transpuesta de A, lo que implica que $A^* = A^\top \in \mathbb{M}_{m\times n}(\mathbb{R})$.

Proposición 5.5. *Sean* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *e* $(\mathbf{Y}, \langle\cdot,\cdot\rangle)$ *dos espacios de Hilbert. Considere* $A : \mathbf{X} \to \mathbf{Y}$ *un operador lineal continuo y* $f : \mathbf{Y} \to \mathbb{R} \cup \{+\infty\}$ *una función propia, convexa y s.c.i.. Suponga que* f *es continua en algún* $y_0 \in im(A)$*, entonces se tiene que*

$$\partial(f \circ A)(x) = A^* \partial f(Ax), \quad \forall x \in \mathbf{X}.$$

Demostración. Tal como con el Teorema de Moreau-Rockafellar, una de las inclusiones es fácil y la otra requiere más desarrollo. Sea $x \in \mathbf{X}$ y comencemos con la inclusión $A^*\partial f(Ax) \subseteq \partial(f \circ A)(x)$, que es la más directa. De la definición misma se tiene que si $y^* \in \partial f(Ax)$ entonces

$$f(Ax) + \langle y^*, y - Ax \rangle \leq f(y), \quad \forall y \in \mathbf{Y}.$$

En particular, esto es cierto para $y = Az$, con $z \in \mathbf{X}$ arbitrario. Entonces, usando la definición de operador adjunto tenemos que $A^*y^* \in \partial(f \circ A)(x)$, pues

$$f(Ax) + \langle A^*y^*, z - x \rangle \leq f(Az), \quad \forall z \in \mathbf{X}.$$

Veamos ahora la otra inclusión. Dado que f es continua en algún $y_0 \in \operatorname{im}(A)$, tenemos que $\operatorname{int}(\operatorname{epi}(f)) \neq \emptyset$ y además la siguiente inclusión siempre es cierta

$$\operatorname{int}(\operatorname{epi}(f)) \subseteq \{(y, \lambda) \in \mathbf{Y} \times \mathbb{R} \mid \ f(y) < \lambda\}.$$

Por lo tanto, dados $x \in \mathbf{X}$ y $x^* \in \partial(f \circ A)(x)$, el conjunto

$$B = \{(Az, f(Ax) + \langle x^*, z - x \rangle) \in \mathbf{Y} \times \mathbb{R} \mid \ z \in \mathbf{X}\}$$

puede ser separado del conjunto $\operatorname{int}(\operatorname{epi}(f))$. Efectivamente, ambos conjuntos son convexos (Proposición 3.1) y no vacíos, y además, si $(y, \lambda) \in B \cap \operatorname{int}(\operatorname{epi}(f))$, entonces para algún $z \in \mathbf{X}$ debemos tener que $y = Az$, $\lambda = f(Ax) + \langle x^*, z - x \rangle$ y

$$f(Az) = f(y) < \lambda = f(Ax) + \langle x^*, z - x \rangle.$$

Pero esta desigualdad estricta es imposible ya que $x^* \in \partial(f \circ A)(x)$. Por lo tanto, $B \cap \operatorname{int}(\operatorname{dom}(f)) = \emptyset$ y por el Teorema de separación de de Hahn-Banach (Lema 3.1), existe $(y^*, r) \in \mathbf{Y}^* \times \mathbb{R} \setminus \{0\}$ tal que

$$\langle y^*, y \rangle + r\lambda < \langle y^*, Az \rangle + r(f(Ax) + \langle x^*, z - x \rangle), \qquad \forall (y, \lambda) \in \operatorname{int}(\operatorname{epi}(f)),\ z \in \mathbf{X}. \tag{5.2}$$

Evaluando en $y = Ax$, $z = x$, $\lambda > f(Ax)$, obtenemos que $r < 0$. Normalizando, podemos entonces asumir que $r = -1$. Luego, en virtud de la Proposición 3.1, para todo $y \in \operatorname{dom} f$, existe una sucesión $\{(y_n, \lambda_n)\}_{n \in \mathbb{N}}$ en $\operatorname{int}\operatorname{epi}(f)$ (que satisface (5.2)) tal que $(y_n, \lambda_n) \to (y, f(y)) \in \operatorname{epi}(f)$, de donde haciendo $n \to \infty$ se obtiene

$$\langle y^*, y \rangle - f(y) \leq \langle y^*, Az \rangle - f(Ax) - \langle x^*, z - x \rangle, \qquad \forall y \in \operatorname{dom} f,\ z \in \mathbf{X}. \tag{5.3}$$

Por otra parte, evaluando (5.3) en $y = Ax$ y $z = x \pm d$ para algún $d \in \mathbf{X} \setminus \{0\}$, llegamos a

$$\langle A^*y^* - x^*, d \rangle = 0, \quad \forall d \in \mathbf{X} \setminus \{0\}$$

de donde podemos concluir que $x^* = A^*y^*$. Ahora bien, evaluando (5.3) en $z = x$ obtenemos

$$f(Ax) + \langle y^*, y - Ax \rangle \leq f(y), \quad \forall y \in \operatorname{dom}(f),$$

lo que implica que $y^* \in \partial f(Ax)$ y por lo tanto $x^* \in A^*\partial f(Ax)$, lo que completa la demostración. □

El siguiente es un contraejemplo que muestra que la igualdad no se tiene si la condición que la función sea continua en un punto de la imagen de A.

Ejemplo 5.1.4. *Supongamos que* $\mathbf{X} = \mathbb{R}$, $\mathbf{Y} = \mathbb{R}^2$, *sea* $C = \{(x,y) \in \mathbb{R}^2 \,|\, x^2 + (y-1)^2 \leq 1\}$, $f = \delta_C$ *y* $A\colon x \mapsto (x, 0)$. *Luego* $\operatorname{im} A = \mathbb{R} \times \{0\}$, $A^*\colon (x,y) \mapsto x$, $\operatorname{dom} f \cap \operatorname{im} A = C \cap (\mathbb{R} \times \{0\}) = \{(0,0)\}$, *de donde* $f \circ A = \delta_{\{0\}}$ *y* $\partial(f \circ A)(0) = \mathbb{R}$. *Por otra parte,* $A0 = (0,0)$ *y* $\partial f(A0) = \partial\delta_C(0,0) = \{0\} \times \mathbb{R}_-$, *de donde* $A^*\partial f(A0) = \{0\} \subsetneq \mathbb{R} = \partial(f \circ A)(0)$. *Notar que* f *no es continua en* $\{(0,0)\} = \operatorname{dom} f \cap \operatorname{im} A$.

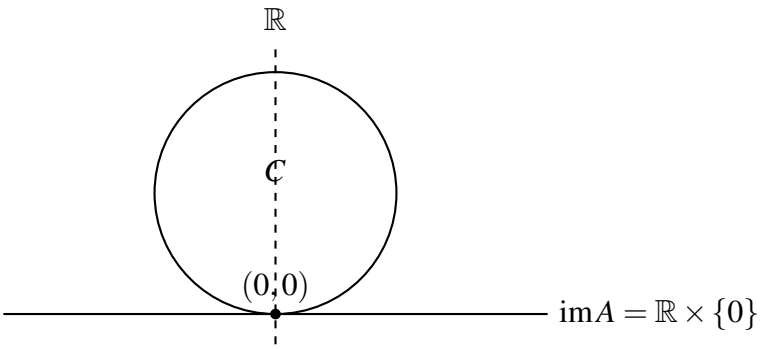


Figura 5.4: Contraejemplo regla de la composición.

5.2. Condiciones de optimalidad

Volvamos ahora al problema

(P) Minimizar $f(x)$ sobre todos los $x \in \mathbf{X}$ que satisfacen la restricción $x \in \mathbf{S}$.

Los resultados anteriores nos proveen las herramientas suficientes para poder ahora escribir las condiciones de optimalidad para este problema. Estas condiciones se escribirán en términos del subdiferencial de la función objetivo y el cono normal al conjunto de restricciones.

Teorema 5.3 (Regla de Fermat III)**.** *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbert,* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función propia convexa y s.c.i., y* $\mathbf{S} \subseteq \mathbf{X}$ *un conjunto convexo cerrado y no vacío. Supongamos que alguna de las siguientes condiciones es cierta:*

1. *Existe* $x_0 \in int(\mathbf{S})$ *tal que* f *es finita en* x_0*.*

2. *Existe* $x_0 \in \mathbf{S}$ *tal que* f *es continua en* x_0*.*

Entonces, $\bar{x} \in \mathbf{S}$ *es una solución de* (P) *si y sólo si*

$$0 \in \partial f(\bar{x}) + N_{\mathbf{S}}(\bar{x})$$

o equivalentemente

$$\exists x^* \in \partial f(\bar{x}) \quad tal\ que\ \langle x^*, x - \bar{x} \rangle \geq 0, \quad \forall x \in \mathbf{S}.$$

Demostración. Notemos que $\bar{x}$ es una solución de (P) si y sólo si $\bar{x} \in \arg\min(f_{\mathbf{S}})$, con $f_{\mathbf{S}} = f + \delta_{\mathbf{S}}$. Luego, por la Regla de Fermat (Teorema 5.1), $\bar{x}$ es una solución de (P) si y sólo si $0 \in \partial f_{\mathbf{S}}(\bar{x})$. Finalmente, cualquiera de las condiciones de calificación del enunciado implican la hipótesis del Teorema de Moreau-Rockafellar (Teorema 5.2). Por lo tanto, aplicando ese resultado obtenemos el resultado buscado pues

$$\partial f_{\mathbf{S}}(\bar{x}) = \partial f(\bar{x}) + \partial \delta_{\mathbf{S}}(\bar{x}) = \partial f(\bar{x}) + N_{\mathbf{S}}(\bar{x}).$$

□

5.2.1. Aplicación a la Programación Convexa

Estudiaremos ahora un problema particular en optimización el cual es conocido como problema de *programación convexa* y que consiste en minimizar una función convexa $f : \mathbf{X} \to \mathbb{R}$ sobre el conjunto de restricciones

$$\mathbf{S} = \left\{x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1,\dots,p, \quad \langle x_j^*, x\rangle = \alpha_j,\ j = 1,\dots,q\right\}.$$

donde $g_1,\dots,g_p : \mathbf{X} \to \mathbb{R}$ son funciones convexas, $x_1^*,\dots,x_q^* \in \mathbf{X}$ y $\alpha_1,\dots,\alpha_q \in \mathbb{R}$.

Estudiaremos las condiciones de optimalidad de problemas de programación convexa primero para el caso sin restricciones de igualdad, y luego extenderemos el resultado a esa situación.

Restricciones de desigualdad

Concentrémonos en el problema de programación convexa siguiente:

($\mathrm{P_D}$) Minimizar $f(x)$ sobre los $x \in \mathbf{X}$ que satisfacen la restricción $g_i(x) \leq 0$ para $i = 1,\dots,p$.

Para obtener las condiciones de optimalidad del problema precedente necesitamos primero el converso de la Proposición 5.2. Por simplicidad mostraremos el resultado para espacio de Banach reflexivo, sin embargo el resultado es igual de válido para espacio que no lo son.

Proposición 5.6. *Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R}$ una función convexa y continua tal que $\Gamma_\gamma(f) \neq \emptyset$ para cierto $\gamma > \text{ínf}_{\mathbf{X}}(f)$. Luego para todo $x \in \Gamma_\gamma(f)$ se tiene*

$$\eta \in N_{\Gamma_\gamma(f)}(x) \quad \Longrightarrow \quad \exists \mu \geq 0,\ \exists x^* \in \partial f(x),\ \textit{tales que } \eta = \mu x^* \textit{ y } \mu(f(x) - \gamma) = 0.$$

Demostración. Recordemos primero que, gracias a Proposición 5.1, tenemos que $\partial f(x) \neq \emptyset$ para todo $x \in \mathbf{X}$, pues f es continua en $\mathbf{X}$. Separemos la demostración en varias etapas.

1. Tomemos $\eta \in N_{\Gamma_\gamma(f)}(x)$ y asumamos que $\eta \neq 0$; el caso $\eta = 0$ es directo de tomar $\mu = 0$. Luego, por definición tenemos
$$\langle \eta, y\rangle \leq \langle \eta, x\rangle, \quad \forall y \in \Gamma_\gamma(f).$$
Notemos que si $f(y) < \gamma$ entonces necesariamente se tendrá que $\langle \eta, y\rangle < \langle \eta, x\rangle$. En efecto, si esto no es así, dado que $\eta \in N_{\Gamma_\gamma(f)}(x)$ deberíamos tener que $\langle \eta, y\rangle = \langle \eta, x\rangle$, pero por continuidad de f se tendrá que existe $r > 0$ tal que $\mathbb{B}_{\mathbf{X}}(y, r) \subseteq \Gamma_\gamma(f)$, con lo cual podemos afirmar que
$$\langle \eta, y + rd\rangle \leq \langle \eta, x\rangle, \quad \forall d \in \mathbb{B}_{\mathbf{X}},$$
y por lo tanto, dado que $\langle \eta, y\rangle = \langle \eta, x\rangle$, tenemos
$$r\langle \eta, d\rangle \leq 0, \quad \forall d \in \mathbb{B}_{\mathbf{X}},$$
lo que implica que $\|\eta\|_* = 0$, es decir $\eta = 0$, lo que no puede ser. Notemos además que lo anterior es también válido si $y = x$. En consecuencia, si $f(x) < \gamma$ se tendrá necesariamente que $\eta = 0$, y la conclusión es válida tomando $\mu = 0$.

2. Resta ver ahora el caso $f(x) = \gamma$ para concluir la demostración. Consideremos el conjunto

$$\mathbf{S}_\eta = \{y \in \mathbf{X} \mid \langle \eta, y\rangle \geq \langle \eta, x\rangle\}.$$

Notemos que si $y \in \mathbf{S}_\eta$, entonces usando la contra-recíproca de la afirmación demostrada anteriormente tenemos que $f(y) \geq \gamma = f(x)$. En otras palabras, $x \in \mathbf{X}$ es óptimo del problema

$$\text{Minimizar } f(y) \text{ sobre todos los } y \in \mathbf{X} \text{ que satisfacen la restricción } y \in \mathbf{S}_\eta.$$

Este problema es convexo y por lo tanto gracias al Teorema 5.3 tenemos que

$$0 \in \partial f(x) + N_{\mathbf{S}_\eta}(x).$$

3. Notemos que para cada $\nu \in N_{\mathbf{S}_\eta}(x) \setminus \{0\}$ tenemos que si $\langle \nu, y\rangle = 0$ entonces $\langle \eta, y\rangle = 0$. En efecto, razonando por contradicción si existiese $y \in \mathbf{X}$ tal que $\langle \nu, y\rangle = 0$ pero $\langle \eta, y\rangle \neq 0$, podemos asumir sin pérdida de generalidad que $\langle \eta, y\rangle > 0$. La continuidad implica que podemos encontrar $r > 0$ tal que $\langle \eta, y + rd\rangle \geq 0$ para todo $d \in \mathbb{B}_{\mathbf{X}}$. En particular, tendremos que $y + rd + x \in \mathbf{S}_\eta$. Ahora, dado que $\nu \in N_{\mathbf{S}_\eta}(x)$ y $\langle \nu, y\rangle = 0$ tenemos que

$$r\langle \nu, d\rangle = \langle \nu, y + rd + x - x\rangle \leq 0, \quad \forall d \in \mathbb{B}_{\mathbf{X}}.$$

Esto nos llevaría a concluir que $\|\nu\| = 0$, lo cual no puede ser.

4. Sea $\nu \in N_{\mathbf{S}_\eta}(x) \setminus \{0\}$ y consideremos $B = \{t\nu \mid \ t \in \mathbb{R}\}$. Tenemos entonces que $\eta \in B$, pues si no, por el Teorema de Hahn-Banach (Lema 3.1) existirá $y \in \mathbf{X}$ tal que

$$\langle y, \eta\rangle < \langle y, t\nu\rangle, \quad \forall t \in \mathbb{R}.$$

Como lo anterior es cierto para todo $t \in \mathbb{R}$, necesariamente tenemos que tener $\langle y, \nu\rangle = 0$ y por lo tanto $\langle y, \eta\rangle < 0$. Sigue que $\langle \nu, y\rangle = 0$ y $\langle \eta, y\rangle < 0$, lo cual contradice lo demostrado en el punto anterior.

5. Juntando toda la información anterior llegamos a que existe $\mu \in \mathbb{R}$ y $x^* \in \partial f(x)$ tales que $\eta = \mu x^*$ para el caso $f(x) = \gamma$. Dado que $\eta \neq 0$, entonces $x^* \neq 0$ y $\mu \neq 0$. En particular, por la Regla de Fermat (Teorema 4.4) se tiene que $\gamma = f(x) > \text{ínf}_{\mathbf{X}}(f)$ y por lo tanto existe $y \in \mathbf{X}$ tal que $f(y) < \gamma$. Notemos además que

$$\langle x^*, y - x\rangle \leq f(y) - f(x) = f(y) - \gamma < 0.$$

Finalmente, como $\mu x^* = \eta \in N_{\Gamma_\gamma(f)}(x)$ e $y \in \Gamma_\gamma(f)$ entonces

$$\mu\langle x^*, y - x\rangle = \langle \mu x^*, y - x\rangle \leq 0,$$

de donde se concluye que $\mu \geq 0$, lo cual completa la demostración.

□

Teorema 5.4 (Teorema de Kuhn-Tucker I)**.** *Sean* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert,* $f : \mathbf{X} \to \mathbb{R}$ *y* $g_1, \ldots, g_p : \mathbf{X} \to \mathbb{R}$ *funciones convexas y continuas. Supongamos que existe* $x_0 \in \mathbf{X}$ *tal que*

$$g_i(x_0) < 0, \quad \forall i = 1, \ldots, p.$$

Entonces, $\bar{x} \in \mathbf{X}$ *es una solución de* (P_D) *si y sólo si existen* $\mu_1, \ldots, \mu_p \geq 0$ *tales que*

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{p} \mu_i \partial g_i(\bar{x}) \tag{5.4}$$

$$g_i(\bar{x}) \leq 0 \quad y \quad \mu_i g_i(\bar{x}) = 0, \quad \forall i = 1, \ldots, p. \tag{5.5}$$

Demostración. Notemos que gracias al Teorema 5.3, $\bar{x} \in \mathbf{X}$ es una solución de (P_D) si y sólo si

$$0 \in \partial f(\bar{x}) + N_{\mathbf{S}}(\bar{x}),$$

con $\mathbf{S} = \{x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1, \ldots, p\}$. Recordemos que $N_{\mathbf{S}}(x) = \partial \delta_{\mathbf{S}}(x)$ para cada $x \in \mathbf{X}$. Además, si denotamos $\mathbf{S}_i = \{x \in \mathbf{X} \mid g_i(x) \leq 0\}$ para cada $i \in \{1, \ldots, p$, sigue que

$$\delta_{\mathbf{S}}(x) = \sum_{i=1}^{p} \delta_{\mathbf{S}_i}(x), \quad \forall x \in \mathbf{X}.$$

Notemos que dado que existe $x_0 \in \mathbf{X}$ tal que

$$g_i(x_0) < 0, \quad \forall i = 1, \ldots, p$$

podemos aplicar recursivamente la regla de la suma para el subdiferencial, y obtener

$$N_{\mathbf{S}}(x) = \partial \delta_{\mathbf{S}}(x) = \sum_{i=1}^{p} \partial \delta_{\mathbf{S}_i}(x) = \sum_{i=1}^{p} N_{\mathbf{S}_i}(x), \quad \forall x \in \mathbf{X}.$$

Para concluir resta ver que para cada $x \in \mathbf{S}_i$ se tiene que

$$\eta \in N_{\mathbf{S}_i}(x) \iff \exists \mu_i \geq 0,\ \exists x_i^* \in \partial g_i(x),\ \text{tales que } \eta = \mu_i x_i^* \text{ y } \mu_i g_i(x) = 0.$$

Pero esto es consecuencia directa de la Proposición 5.2 y Proposición 5.6. Luego el teorema ha sido demostrado. □

Restricciones de desigualdad e igualdad

Retomemos el problema general de programación convexa, es decir,

(P_{DI}) Minimizar $f(x)$ sobre los $x \in \mathbf{X}$ tales que $g_i(x) \leq 0$ para $i = 1, \ldots, p$ y $\ell(x) = (\alpha_1, \ldots, \alpha_q)$

donde $\ell : \mathbf{X} \to \mathbb{R}^q$ es el funcional lineal continuo dado por

$$\ell(x) = \left(\langle x_1^*, x\rangle, \ldots, \langle x_q^*, x\rangle\right), \qquad \forall x \in \mathbf{X}.$$

La versión que estudiaremos ahora del Teorema de Kuhn-Tucker es una extensión de Teorema 5.4.

Teorema 5.5 (Teorema de Kuhn-Tucker II)**.** *Sean* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert,* $f : \mathbf{X} \to \mathbb{R}$ *y* $g_1, \ldots, g_p : \mathbf{X} \to \mathbb{R}$ *funciones convexas y continuas. Sean* $x_1^*, \ldots, x_q^* \in \mathbf{X}$ *y* $\alpha_1, \ldots, \alpha_q \in \mathbb{R}$ *dados. Supongamos que existe* $x_0 \in \mathbf{X}$ *tal que*

$$g_i(x_0) < 0, \quad \forall i = 1, \ldots, p. \quad y \quad \langle x_j^*, x_0\rangle = \alpha_j, \quad \forall j = 1, \ldots, q$$

Entonces, $\bar{x} \in \mathbf{X}$ *es una solución de* (P_{DI}) *si y sólo si existen* $\mu_1, \ldots, \mu_p \geq 0$ *y* $\lambda_1, \ldots, \lambda_q \in \mathbb{R}$ *tales que*

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{p} \mu_i \partial g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j x_j^* \tag{5.6}$$

$$g_i(\bar{x}) \leq 0 \quad y \quad \mu_i g_i(\bar{x}) = 0, \quad \forall i = 1, \ldots, p. \tag{5.7}$$

$$\langle x_j^*, \bar{x}\rangle = \alpha_j, \quad \forall j = 1, \ldots, q. \tag{5.8}$$

Demostración. Notemos que gracias al Teorema 5.3, $\bar{x} \in \mathbf{X}$ es una solución de (P_{DI}) si y sólo si

$$0 \in \partial f(\bar{x}) + N_{\mathbf{S}}(\bar{x}) + N_H(\bar{x}),$$

con $\mathbf{S} = \{x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1, \ldots, p\}$ y $H = \left\{x \in \mathbf{X} \mid \langle x_j^*, x\rangle = \alpha_j,\ j = 1, \ldots, q\right\}$. Ya hemos visto que

$$\eta \in N_{\mathbf{S}}(\bar{x}) \quad \Longleftrightarrow \quad \exists \mu_1, \ldots, \mu_p \geq 0, \text{ tales que } \mu_i g_i(\bar{x}) = 0 \text{ y } \eta \in \sum_{i=1}^{p} \mu_i \partial g_i(\bar{x}).$$

Luego, para concluir basta ver que

$$\eta \in N_H(\bar{x}) \quad \Longleftrightarrow \quad \exists \lambda_1, \ldots, \lambda_q \in \mathbb{R}, \text{ tales que } \eta = \sum_{j=1}^{q} \lambda_j x_j^*.$$

Dividamos la demostración de esta equivalencia en partes:

1. No es difícil ver que para cualquier $\lambda_1, \ldots, \lambda_q \in \mathbb{R}$ se tiene

$$\left\langle \sum_{j=1}^{q} \lambda_j x_j^*, x - \bar{x} \right\rangle = \sum_{j=1}^{q} \langle \lambda_j x_j^*, x - \bar{x}\rangle = \sum_{j=1}^{q} \lambda_j \left(\langle x_j^*, x\rangle - \langle x_j^*, \bar{x}\rangle\right) = \sum_{j=1}^{q} \lambda_j \left(\langle x_j^*, x\rangle - \alpha_j\right).$$

   Lo que a su vez implica que $\sum_{j=1}^{q} \lambda_j x_j^* \in N_H(\bar{x})$. Hemos aquí demostrado la implicancia ($\Leftarrow$).

2. Veamos ahora que $N_H(\bar{x})$ es un espacio vectorial. Dado que $N_H(\bar{x})$ es un conjunto convexo, bastará mostrar que si $\eta \in N_H(\bar{x})$ entonces $-\eta \in N_H(\bar{x})$. En efecto, notemos que si $\eta \in N_H(\bar{x})$ entonces

$$\langle \eta, x - \bar{x}\rangle \leq 0, \quad \forall x \in H$$

   y que además, si $x \in H$, entonces igualmente $2\bar{x} - x \in H$. Esto último se debe a que

$$\langle x_j^*, 2\bar{x} - x\rangle = 2\alpha_j - \alpha_j = \alpha_j, \qquad \forall j = 1, \ldots, q.$$

   Entonces tenemos que

$$\langle -\eta, x - \bar{x}\rangle = \langle \eta, \bar{x} - x\rangle = \langle \eta, (2\bar{x} - x) - \bar{x}\rangle \leq 0, \quad \forall x \in H.$$

3. Notemos que lo demostrado en el paso anterior implica que para todo $x \in \mathbf{X}$ la siguiente propiedad es cierta

$$\langle x_j^*, x-\bar{x}\rangle = 0, \quad \forall j=1,\dots,q \qquad \Longrightarrow \qquad \langle \eta, x-\bar{x}\rangle = 0, \quad \forall \eta \in N_H(\bar{x}). \tag{5.9}$$

Consideremos el espacio vectorial

$$B = \left\{ x^* \in \mathbf{X} \,\middle|\, \exists \lambda_1,\dots,\lambda_q \in \mathbb{R} \text{ tal que } x^* = \sum_{j=1}^{q} \lambda_j x_j^* \right\}.$$

Queremos demostrar que cualquier $\eta \in N_H(\bar{x})$ pertenece a B. Supongamos por contradicción que esto no es así. Luego por el Teorema de Separación de Hahn-Banach (Lema 3.1) podemos separar estrictamente η del conjunto B,i.e., existe $x \in \mathbf{X} \setminus \{\bar{x}\}$ tal que

$$\langle \eta, x-\bar{x}\rangle < \sum_{j=1}^{q} \lambda_j \langle x_j^*, x-\bar{x}\rangle, \quad \forall \lambda_1,\dots,\lambda_q \in \mathbb{R}.$$

Esto implica que $\langle x_j^*, x-\bar{x}\rangle = 0$ para todo $j=1,\dots,q$ pues si no, podemos hacer $\lambda_j \to \pm\infty$ y llegar a una contradicción. Ahora bien, por (5.9) tenemos que $\langle \eta, x-\bar{x}\rangle = 0$, lo cual tampoco puede ser. Por lo tanto, η debe pertenecer a B y la conclusión sigue.

□

Lagrangiano de un problema de programación convexa

Veremos a continuación una lectura diferente del Teorema de Kuhn-Tucker, la cual es una forma equivalente del resultado, pero que sin embargo entrega una visión distinta del problema de programación convexa.

Consideremos la función Lagrangeana asociada al problema de programación convexa general (P_{DI}), que denotamos $L : \mathbf{X} \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$, y que está dada por

$$L(x,\mu,\lambda) := f(x) + \sum_{i=1}^{p} \mu_i g_i(x) + \sum_{j=1}^{q} \lambda_j (\langle x_j^*, x\rangle - \alpha_j), \quad \forall x \in \mathbf{X},\ \mu \in \mathbb{R}^p,\ \lambda \in \mathbb{R}^q.$$

Notemos que para $\mu_1,\dots,\mu_p \geq 0$ y $\lambda_1,\dots,\lambda_q \in \mathbb{R}$ fijos, uno tiene que la función $x \mapsto L(x,\mu,\lambda)$ es convexa. Luego, bajo hipótesis simples podemos esperar que se tenga la igualdad

$$\partial_x L(x,\mu,\lambda) := \partial(L(\cdot,\mu,\lambda))(x) = \partial f(\bar{x}) + \sum_{i=1}^{p} \mu_i \partial g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j x_j^*, \quad \forall x \in \mathbf{X}.$$

Cabe destacar que en el caso diferenciable, tendremos que

$$\partial_x L(x,\mu,\lambda) = \left\{ \nabla f(\bar{x}) + \sum_{i=1}^{p} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j x_j^* \right\}, \quad \forall x \in \mathbf{X}.$$

Sigue que el Teorema de Kuhn-Tucker se puede re-escribir de la siguiente forma:

Teorema 5.6 (Teorema de Kuhn-Tucker III)**.** *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert,* $f : \mathbf{X} \to \mathbb{R}$ *y* $g_1, \ldots, g_p : \mathbf{X} \to \mathbb{R}$ *funciones convexas y continuas. Sean* $x_1^*, \ldots, x_q^* \in \mathbf{X}$ *y* $\alpha_1, \ldots, \alpha_q \in \mathbb{R}$ *dados. Supongamos que existe* $x_0 \in \mathbf{X}$ *tal que*

$$g_i(x_0) < 0, \quad \forall i = 1, \ldots, p. \quad y \quad \langle x_j^*, x_0\rangle = \alpha_j, \quad \forall j = 1, \ldots, q$$

Entonces, $\bar{x} \in \mathbf{X}$ *es una solución de* (P_{DI}) *si y sólo* $\bar{x}$ *es factible, es decir,*

$$g_i(\bar{x}) \leq 0, \quad \forall i = 1, \ldots, p \quad y \quad \langle x_j^*, \bar{x}\rangle = \alpha_j, \quad \forall j = 1, \ldots, q,$$

y además existen $\mu_1, \ldots, \mu_p \geq 0$ *y* $\lambda_1, \ldots, \lambda_q \in \mathbb{R}$ *tales que*

$$0 \in \partial_x L(\bar{x}, \mu, \lambda) \quad y \quad \mu_i g_i(\bar{x}) = 0, \quad \forall i = 1, \ldots, p.$$

Más aún, si la función objetivo f *y las funciones* $g_1, \ldots, g_p$ *son diferenciables en una vecindad de* $\bar{x}$*, entonces la condición anterior es equivalente a*

$$0 = \nabla_x L(\bar{x}, \mu, \lambda) = \nabla f(\bar{x}) + \sum_{i=1}^{p} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j x_j^* \quad y \quad \mu_i g_i(\bar{x}) = 0, \quad \forall i = 1, \ldots, p.$$

5.3. Aproximación de Moreau-Yosida

En esta sección estudiaremos un forma de aproximar funciones convexas no diferenciables, usando una secuencia de funciones convexas que si lo son. Este esquema de aproximación dará paso a introducir métodos numéricos para resolver problemas de optimización convexa no diferenciable.

En adelante nos situaremos en el contexto de espacios de Hilbert, es decir, $\mathbf{X}$ es un espacio de Banach y la norma $\|\cdot\|$ está inducida por un producto interno $\langle\cdot,\cdot\rangle$. Como mencionamos al comienzo, $\mathbf{X}$ será identificado con $\mathbf{X}$ y el producto dualidad será el mismo que el producto interno.

Definición 5.2 (Aproximación de Moreau-Yosida)**.** *Sean* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función dada. Para* $\alpha > 0$*, la* aproximación Moreau-Yosida *de* f *es la función*

$$f_\alpha(x) := \inf_{y \in \mathbf{X}} \left\{ f(y) + \frac{1}{2\alpha}\|x - y\|^2 \right\}, \quad \forall x \in \mathbf{X}.$$

En la Figura 5.5 se muestran dos ejemplos para $f = |\cdot|$ y para $f = \delta_{[-1,1]}$ con $\alpha = 1$.

La siguiente proposición resume las principales características de la aproximación de Moreau-Yosida de funciones convexas.

Proposición 5.7. *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función propia convexa y s.c.i. Si* $\alpha > 0$ *está fijo, entonces* f_α *es convexa, Fréchet diferenciable en* $\mathbf{X}$ *y para todo* $x \in \mathbf{X}$ *existe un único* $y_\alpha(x) \in dom(f)$ *tal que*

$$\nabla f_\alpha(x) = \frac{x - y_\alpha(x)}{\alpha} \in \partial f(y_\alpha(x)) \quad y \quad f_\alpha(x) = f(y_\alpha(x)) + \frac{1}{2\alpha}\|x - y_\alpha(x)\|^2.$$

Además, las aplicaciones $\nabla f_\alpha : \mathbf{X} \to \mathbf{X}$ *e* $y_\alpha : \mathbf{X} \to \mathbf{X}$ *son Lipschitz continuas de constante* $\frac{1}{\alpha}$ *y* 1*, respectivamente. También se tiene que*

$$\lim_{\alpha \to 0} y_\alpha(x) = x \quad y \quad \lim_{\alpha \to 0} f_\alpha(x) = f(x), \qquad \forall x \in \overline{dom(f)}.$$

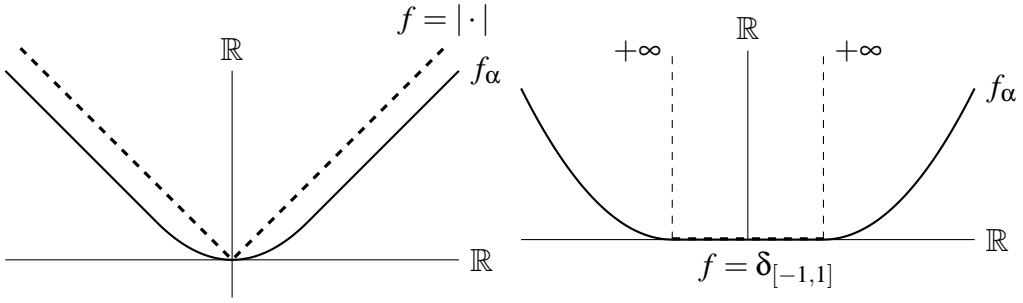


Figura 5.5: Ejemplos de f_α con $\alpha = 1$.

Demostración. Dividiremos la demostración en varias partes:

1. Comencemos mostrando la existencia y unicidad de $y_\alpha(x)$. Notemos que $y_\alpha(x)$ debe ser un mínimo del problema de optimización dado que define a la aproximación de Moreau-Yosida. Más aún, como la función $y \mapsto f(y) + \frac{1}{2\alpha}\|x-y\|^2$ es estrictamente convexa tenemos que ésta tiene a lo más un mínimo, de donde obtenemos la unicidad. Para la existencia veamos que $y \mapsto f(y) + \frac{1}{2\alpha}\|x-y\|^2$ es coerciva. En efecto, como f es convexa propia y s.c.i., gracias a Proposición 3.3, existe $x^* \in \mathbf{X}$ y $c \in \mathbb{R}$ tal que

$$f(y) + \frac{1}{2\alpha}\|x-y\|^2 \geq \langle x^*, y\rangle + c + \frac{1}{2\alpha}\|x-y\|^2 \geq g(\|x-y\|), \quad \forall y \in \mathbf{X},$$

donde $g(t) = \frac{1}{2\alpha}t^2 - \|x^*\|_* t + \langle x^*, x\rangle + c$. Notemos que g es una función cuadrática, entonces por Proposición 4.2 ésta es coerciva, de donde obtenemos que $y \mapsto f(y) + \frac{1}{2\alpha}\|x-y\|^2$ es también coerciva. Luego, gracias al teorema de Weierstrass-Hilbert-Tonelli (Teorema 3.1) podemos asegurar la existencia de $y_\alpha(x)$ para cualquier $x \in \mathbf{X}$. Más aún, gracias a la regla de Fermat y la regla de la suma para subdiferenciales (Teorema de Moreau-Rockafellar), tenemos que

$$0 \in \partial f(y_\alpha(x)) + \frac{y_\alpha(x) - x}{\alpha}, \quad \forall x \in \mathbf{X}.$$

2. Como $(x,y) \mapsto f(y) + \frac{1}{2\alpha}\|x-y\|^2$ es convexa, la convexidad de f_α es directa del Ejercicio 2. Veamos ahora que f_α es Fréchet diferenciable con $\nabla f_\alpha(x) = \frac{x - y_\alpha(x)}{\alpha}$. Por la parte anterior para cualquier $h \in \mathbf{X}$ tenemos que

$$f_\alpha(x+h) - f_\alpha(x) = f(y_\alpha(x+h)) - f(y_\alpha(x)) + \frac{1}{2\alpha}\left(\|x+h-y_\alpha(x+h)\|^2 - \|x-y_\alpha(x)\|^2\right).$$

Dado que $\frac{x-y_\alpha(x)}{\alpha} \in \partial f(y_\alpha(x))$ podemos deducir que

$$f_\alpha(x+h) - f_\alpha(x) \geq \frac{1}{\alpha}\langle x - y_\alpha(x), y_\alpha(x+h) - y_\alpha(x)\rangle + \frac{1}{2\alpha}\left(\|x+h-y_\alpha(x+h)\|^2 - \|x-y_\alpha(x)\|^2\right)$$

pero

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

lo que implica que, si denotamos $x^* = \frac{x-y_\alpha(x)}{\alpha}$, entonces

$$f_\alpha(x+h) - f_\alpha(x) - \langle x^*, h\rangle \geq \|y_\alpha(x+h) - y_\alpha(x) - h\|^2 \geq 0.$$

Por otro lado, por definición de la aproximación de Moreau-Yosida tenemos

$$\begin{aligned} f_\alpha(x+h) - f_\alpha(x) &\leq f(y_\alpha(x)) + \frac{1}{2\alpha}\|x+h-y_\alpha(x)\|^2 - f(y_\alpha(x)) - \frac{1}{2\alpha}\|x-y_\alpha(x)\|^2 \\ &= \frac{1}{2\alpha}\left(\|x+h-y_\alpha(x)\|^2 - \|x-y_\alpha(x)\|^2\right) = \frac{1}{2\alpha}\left(\|h\|^2 + 2\langle x-y_\alpha(x), h\rangle\right) \end{aligned}$$

Esto a su vez nos lleva a concluir que f_α es Fréchet diferenciable con $\nabla f_\alpha(x) = x^* = \frac{x-y_\alpha(x)}{\alpha}$ pues, reuniendo las desigualdades anteriores llegamos a:

$$0 \leq f_\alpha(x+h) - f_\alpha(x) - \langle x^*, h\rangle \leq \frac{1}{2\alpha}\|h\|^2, \quad \forall h \in \mathbf{X}.$$

3. Veamos que y_α no-expansiva. Para ello notemos que $\nabla f_\alpha(x) \in \partial f(y_\alpha(x))$, luego usando la monotonía del subdiferencial tenemos

$$\langle \nabla f_\alpha(x+h) - \nabla f_\alpha(x), y_\alpha(x+h) - y_\alpha(x)\rangle \geq 0, \quad \forall h \in \mathbf{X},$$

pero, esto implica que

$$\langle h - y_\alpha(x+h) + y_\alpha(x), y_\alpha(x+h) - y_\alpha(x)\rangle \geq 0, \quad \forall h \in \mathbf{X},$$

y por lo tanto

$$\|h\|\|y_\alpha(x+h) - y_\alpha(x)\| \geq \langle h, y_\alpha(x+h) - y_\alpha(x)\rangle \geq \|y_\alpha(x+h) - y_\alpha(x)\|^2, \quad \forall h \in \mathbf{X}.$$

Dividiendo por $\|y_\alpha(x+h) - y_\alpha(x)\|$ se obtiene el resultado buscado.

4. El hecho que ∇f_α es Lipschitz continuo viene de la siguiente desigualdad:

$$\begin{aligned} \|\nabla f_\alpha(x+h) - \nabla f_\alpha(x)\|^2 &= \frac{1}{\alpha^2}\|h - y_\alpha(x+h) - y_\alpha(x)\|^2 \\ &= \frac{1}{\alpha^2}\left(\|h\|^2 - 2\langle h, y_\alpha(x+h) - y_\alpha(x)\rangle + \|y_\alpha(x+h) - y_\alpha(x)\|^2\right) \\ &\leq \frac{1}{\alpha^2}\left(\|h\|^2 - \|y_\alpha(x+h) - y_\alpha(x)\|^2\right) \leq \frac{1}{\alpha^2}\|h\|^2 \end{aligned}$$

5. Por definición de la aproximación de Moreau-Yosida tenemos

$$f(y_\alpha(x)) \leq f_\alpha(x) = f(y_\alpha(x)) + \frac{1}{2\alpha}\|x - y_\alpha(x)\|^2 \leq f(x), \quad \forall x \in \mathrm{dom}(f).$$

Recordemos que existe $x^* \in \mathbf{X}$ y $c \in \mathbb{R}$ tales que

$$\langle x^*, y_\alpha(x)\rangle + c \leq f(y_\alpha(x)), \quad \forall x \in \mathrm{dom}(f).$$

Con esto vemos que para cada $x \in \text{dom}(f)$ tenemos

$$\|x - y_\alpha(x)\|^2 \leq 2\alpha\,(f(x) - c + \|x^*\|\|y_\alpha(x)\|)$$

y por lo tanto $\|x - y_\alpha(x)\|$ está uniformemente acotado con respecto a $\alpha > 0$. Luego, a posteriori vemos que $\|x - y_\alpha(x)\| \to 0$ si $\alpha \to 0$ y $x \in \text{dom}(f)$. Finalmente, como f es s.c.i tenemos que

$$f(x) \leq \liminf_{\alpha\to 0} f(y_\alpha(x)) \leq \limsup_{\alpha\to 0} f(y_\alpha(x)) \leq f(x), \quad \forall x \in \text{dom}(f).$$

Usando esto, y el hecho que y_α es Lipschitz continuo, podemos extender las convergencia al caso $x \in \overline{\text{dom}(f)}$, lo que concluye la demostración

□

5.3.1. Método de Punto Proximal

Las propiedades de la aproximación de Moreau-Yosida nos permiten definir, para toda función f propia, convexa y s.c.i. y $\alpha > 0$, el operador $\text{prox}_{\alpha f} : \mathbf{X} \to \mathbf{X}$ como

$$\text{prox}_{\alpha f}(x) := y_\alpha(x), \qquad \forall x \in \mathbf{X},$$

donde $y_\alpha(x)$ está dado por la Proposición 5.7. La existencia del *operador proximal de f de constante* α es una consecuencia de la Proposición 5.7. Notemos también que ese resultado permite caracterizar al operador proximal como la única solución, para $x \in \mathbf{X}$ dado, de la inclusión

$$x \in y + \alpha\partial f(y).$$

Ejemplo 5.3.1. *Sea $\mathbf{S} \subseteq \mathbf{X}$ un conjunto convexo, cerrado y no vacío. Luego, no es difícil ver que la regularizada de Moreau-Yosida de la función $\delta_\mathbf{S}$ es*

$$(\delta_\mathbf{S})_\alpha(x) = \frac{1}{2\alpha} dist^2(x, \mathbf{S}), \quad \forall x \in \mathbf{X}.$$

Aquí $x \mapsto dist(x, \mathbf{S})$ es la función distancia (ver Ejercicio 6 - Capítulo 3) Por lo tanto, $\text{prox}_{\alpha\delta_\mathbf{S}}(x)$ no es otra cosa que la proyección de x sobre $\mathbf{S}$, para todo $\alpha > 0$.

Para aproximar los mínimos de f, proponemos generar una sucesión via la recurrencia

$$x_{k+1} = \text{prox}_{\alpha_k f}(x_k), \quad \forall k \in \mathbb{N}, \tag{5.10}$$

donde la condición inicial $x_0 \in \mathbf{X}$ es arbitraria y $\alpha_k > 0$. En otras palabras tenemos que

$$f_{\alpha_k}(x_k) = f(x_{k+1}) + \frac{1}{2\alpha_k}\|x_k - x_{k+1}\|^2.$$

Estudiaremos ahora la convergencia de una sucesión generada por (5.10), el cual se conoce como *Método de Punto Proximal*.

Teorema 5.7. *Sean* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función propia convexa y s.c.i. tal que* $\arg\min_{\mathbf{X}}(f)$ *es no vacío. Considere una sucesión* $\{\alpha_k\} \subseteq \mathbb{R}$ *que satisface*

$$\inf_{k\in\mathbb{N}} \alpha_k = \alpha > 0$$

y la sucesión $\{x_k\}$ *generada por* (5.10) *partiendo desde* $x_0 \in \mathbf{X}$ *arbitrario. Entonces* $\exists x_\infty \in \arg\min_{\mathbf{X}}(f)$ *tal que* $x_k \rightharpoonup x_\infty$ *cuando* $k \to \infty$.

Demostración. Sea $k \in \mathbb{N}$ y sea $\bar{x} \in \arg\min_{\mathbf{X}}(f)$. Usando la Proposición 5.7 y (5.10) deducimos

$$\frac{x_k - x_{k+1}}{\alpha_k} \in \partial f(x_{k+1}),$$

de donde, por convexidad de f obtenemos

$$f(x_{k+1}) + \frac{1}{\alpha_k}\langle x_k - x_{k+1}, \bar{x} - x_{k+1}\rangle \leq f(\bar{x}) \tag{5.11}$$

o, equivalentemente,

$$f(x_{k+1}) + \frac{1}{2\alpha_k}\left(\|x_k - x_{k+1}\|^2 + \|x_{k+1} - \bar{x}\|^2 - \|x_k - \bar{x}\|^2\right) \leq f(\bar{x}).$$

Usando que $f(\bar{x}) \leq f(y)$ para cualquier $y \in \mathbf{X}$, se obtiene

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 - \|x_k - x_{k+1}\|^2,$$

de donde la sucesión $\{\|x_k - \bar{x}\|\}$ es decreciente y positiva, por lo tanto convergente y $\{x_k\}$ es acotada. Además, sumando sobre k entre 0 y n en la desigualdad anterior y usando la propiedad telescópica deducimos

$$\sum_{k=0}^{n} \|x_k - x_{k+1}\|^2 \leq \|x_0 - \bar{x}\|^2 - \|x_{n+1} - \bar{x}\|^2,$$

de donde concluimos que la serie $\sum_{k=0}^{\infty} \|x_k - x_{k+1}\|^2$ es convergente y luego $x_k - x_{k+1} \to 0$. Para concluir, basta usar el Lema 4.1. Sea $z \in \mathbf{X}$ un punto de acumulación débil de la sucesión $\{x_k\}$, cuya existencia está garantizada por el acotamiento de la misma. Digamos $x_{k_n} \rightharpoonup z$. Usando que f es s.c.i. para la topología débil dado que es convexa (Proposición 3.3) y (5.11) se deduce

$$f(z) \leq \liminf_{k\to+\infty} f(x_{k_n}) = \liminf_{n\to+\infty}\left(f(x_{k_n}) + \frac{1}{\alpha_{k_n-1}}\langle x_{k_n-1} - x_{k_n}, \bar{x} - x_{k_n}\rangle\right) \leq f(\bar{x}),$$

donde la igualdad se obtiene del hecho que

$$\inf_{n\geq 0} \alpha_{k_n} \geq \alpha > 0, \quad x_{k_n-1} - x_{k_n} \to 0 \quad \text{y} \quad x_{k_n} \rightharpoonup z.$$

De ese modo, $z \in \arg\min_{\mathbf{X}}(f)$ y el resultado se deduce de Lema 4.1 con $\mathbf{S} = \arg\min_{\mathbf{X}}(f)$. □

5.4. Método del Gradiente Proximal

Varios de los problemas mencionados en la Sección 1.1 se pueden formular como casos particular del problema de optimización

$$\min_{x\in\mathbf{X}} f(x)+g(x), \tag{5.12}$$

donde $f:\mathbf{X}\to\mathbb{R}\cup\{+\infty\}$ es una función convexa propia s.c.i. y $g\colon \mathbf{X}\to\mathbb{R}$ es otra función convexa, pero Gâteaux diferenciable con gradiente L-Lipschitz continuo. Nos interesa ahora estudiar un método numérico para resolver problemas con esta estructura. El algoritmo que introduciremos se basa en la siguiente idea:

Supongamos que $\bar{x}\in\arg\min_{\mathbf{X}}(f+g)$. Entonces del teorema de Fermat y Teorema de Moreau-Rockafellar (Teorema 5.2) se concluye

$$\bar{x}\in\arg\min_{\mathbf{X}}(f+g) \quad\Leftrightarrow\quad 0\in\partial(f+g)(\bar{x})=\partial f(\bar{x})+\{\nabla g(\bar{x})\}.$$

Notemos que, para todo $\alpha>0$, la condición de optimalidad anterior es equivalente a

$$\bar{x}-\alpha\nabla g(\bar{x})\in\bar{x}+\alpha\partial f(\bar{x}) \quad\Leftrightarrow\quad \bar{x}=\operatorname{prox}_{\alpha f}(\bar{x}-\alpha\nabla g(\bar{x})).$$

Esto motiva el *Método del Gradiente Proximal.* , que está definido a través de la recurrencia

$$x_{k+1}=\operatorname{prox}_{\alpha_k f}(x_k-\alpha_k\nabla g(x_k)),\quad \forall k\in\mathbb{N}, \tag{5.13}$$

donde $x_0\in\mathbf{X}$ es arbitrario y $\alpha_k>0$. Notemos que esta es una extensión natural del método de punto proximal. En efecto, ese algoritmo se recupera si tomamos el caso $g\equiv 0$.

Ahora estudiaremos la convergencia del método del Gradiente Proximal.

Teorema 5.8. *Sea $(\mathbf{X},\langle\cdot,\cdot\rangle)$ un espacio de Hilbert, $f:\mathbf{X}\to\mathbb{R}\cup\{+\infty\}$ y $g:\mathbf{X}\to\mathbb{R}$ dos funciones propias convexas y s.c.i. tal que $\arg\min_{\mathbf{X}}(f+g)$ es no vacío. Supongamos que g es Gâteaux diferenciable en $\mathbf{X}$ con ∇g siendo L-Lipschitz continuo en $\mathbf{X}$. Consideremos $x_0\in\mathbf{X}$ arbitrario, $\varepsilon\in(0,\frac{1}{L})$ y una sucesión $\{\alpha_k\}\subseteq\mathbb{R}$ tal que*

$$\varepsilon\le\alpha_k\le\frac{2}{L}-\varepsilon,\quad\forall k\in\mathbb{N}.$$

Entonces la sucesión $\{x_k\}$ generada por (5.13) converge débilmente a algún $x_\infty\in\arg\min_{\mathbf{X}}(f+g)$.

Demostración. Sea $k\in\mathbb{N}$ y sea $\bar{x}\in\arg\min_{\mathbf{X}}(f+g)$. Usando la Proposición 5.7 y (5.13) deducimos

$$\frac{x_k-x_{k+1}}{\alpha_k}-\nabla g(x_k)\in\partial f(x_{k+1}),$$

de donde, por convexidad de f se obtiene

$$f(x_{k+1})+\frac{1}{\alpha_k}\langle x_k-x_{k+1},y-x_{k+1}\rangle-\langle\nabla g(x_k),y-x_{k+1}\rangle\le f(y),\quad\forall y\in\mathbf{X},$$

o, equivalentemente,

$$f(x_{k+1})\le f(y)+\langle\nabla g(x_k),y-x_{k+1}\rangle+\frac{1}{2\alpha_k}\left(\|x_k-y\|^2-\|x_k-x_{k+1}\|^2-\|x_{k+1}-y\|^2\right),\quad\forall y\in\mathbf{X}.$$

Por otra parte, del Lema 4.2 se obtiene

$$g(x_{k+1}) \leq g(y) + \langle \nabla g(x_k), x_{k+1} - y \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2, \quad \forall y \in \mathbf{X}.$$

Sumando las dos últimas desigualdades se deduce que, para todo $y \in \mathbf{X}$,

(5.14)
$$(f+g)(x_{k+1}) \leq (f+g)(y) + \frac{1}{2\alpha_k}\left(\|x_k - y\|^2 - \|x_k - x_{k+1}\|^2 - \|x_{k+1} - y\|^2\right) + \frac{L}{2}\|x_k - x_{k+1}\|^2.$$

En particular, si tomamos $y = x_k$ obtenemos de $\alpha_k \leq 2/L - \varepsilon$

$$(f+g)(x_{k+1}) \leq (f+g)(x_k) - \left(\frac{1}{\alpha_k} - \frac{L}{2}\right)\|x_k - x_{k+1}\|^2 \leq (f+g)(x_k) - \frac{\varepsilon L^2}{4}\|x_k - x_{k+1}\|^2.$$

Deducimos que la sucesión $\{(f+g)(x_k)\}$ es decreciente y acotada inferiormente por $(f+g)(x^*) = \text{mín}(f+g)$, por lo que converge. Además, sumando sobre k entre 0 y n en la desigualdad anterior y usando la propiedad telescópica deducimos

$$\frac{\varepsilon L^2}{4}\sum_{k=0}^{n}\|x_k - x_{k+1}\|^2 \leq (f+g)(x_0) - (f+g)(x_{n+1})$$

de donde concluimos que la serie $\sum_{k=0}^{\infty}\|x_k - x_{k+1}\|^2$ es convergente y luego $x_k - x_{k+1} \to 0$.

Ahora, tomando $y = \bar{x}$ en (5.14) de $\varepsilon \leq \alpha_k < 2/L$ se tiene

$$\begin{aligned}(f+g)(x_{k+1}) &\leq (f+g)(\bar{x}) + \frac{1}{2\alpha_k}\left(\|x_k - \bar{x}\|^2 - \|x_{k+1} - \bar{x}\|^2 + (\alpha_k L - 1)\|x_k - x_{k+1}\|^2\right)\\ &\leq (f+g)(\bar{x}) + \frac{1}{2\varepsilon}\left(\|x_k - \bar{x}\|^2 - \|x_{k+1} - \bar{x}\|^2 + \|x_k - x_{k+1}\|^2\right), \qquad (5.15)\end{aligned}$$

de donde, usando que $(f+g)(\bar{x}) \leq (f+g)(y)$ para cualquier $y \in \mathbf{X}$, concluimos

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 + \|x_k - x_{k+1}\|^2$$

por lo que, dado que la serie $\sum_{k=0}^{\infty}\|x_k - x_{k+1}\|^2$ converge, se deduce que $\{\|x_{k+1} - \bar{x}\|^2\}$ converge.

Para concluir, basta usar el Lema 4.1. Sea $z \in \mathbf{X}$ un punto de acumulación débil de la sucesión $\{x_k\}$, cuya existencia está garantizada por el acotamiento de la misma. Digamos $x_{k_n} \rightharpoonup z$. Usando (5.15), la semicontinuidad inferior de $f+g$, $x_k - x_{k+1} \to 0$ y que $\{\|x_k - \bar{x}\|^2\}$ converge, deducimos que

$$\begin{aligned}(f+g)(z) &\leq \liminf_{n\to+\infty}(f+g)(x_{k_n})\\ &\leq \liminf_{n\to+\infty}\left((f+g)(\bar{x}) + \frac{1}{2\varepsilon}\left(\|x_{k_n-1} - \bar{x}\|^2 - \|x_{k_n} - \bar{x}\|^2 + \|x_{k_n-1} - x_{k_n}\|^2\right)\right)\\ &= (f+g)(\bar{x}),\end{aligned}$$

de donde $z \in \arg\text{mín}_{\mathbf{X}}(f+g)$ y el resultado se deduce de Lema 4.1 con $\mathbf{S} = \arg\text{mín}_{\mathbf{X}}(f+g)$. □

5.5. Ejercicios

1. CARACTERIZACIÓN DE FUNCIONES CONVEXAS NO DIFERENCIABLE

   Muestre que, análogamente al Teorema 4.1, se tiene que si $(\mathbf{X}, \|\cdot\|)$ un espacio vectorial normado y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ es propia se tiene que las siguientes afirmaciones son equivalentes:

   (i) $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ es convexa.

   (ii) f es subdiferenciable: $f(x) + \langle x^*, y - x\rangle \leq f(y), \quad \forall x, y \in \mathrm{dom}(f),\ x^* \in \partial f(x)$.

   (iii) ∂f es monótono: $\langle x^* - y^*, x - y\rangle \geq 0 \quad \forall x, y \in \mathrm{dom}(f),\ x^* \in \partial f(x),\ y^* \in \partial f(y)$.

2. CONJUGADA DE FENCHEL

   Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función propia convexas y s.c.i. Definimos la función conjugada de f, denotada $f^* : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ via la fórmula:

   $$f^*(x^*) := \sup_{x \in \mathbf{X}} \{\langle x^*, x\rangle - f(x)\}, \quad \forall x^* \in \mathbf{X}.$$

   a) Demuestre que f^* es una función convexa y s.c.i., y que f^* es propia si $\text{ínf}_{\mathbf{X}}(f) > -\infty$.

   b) Pruebe que $x^* \in \partial f(x)$ si y sólo si $f(x) + f^*(x^*) = \langle x^*, x\rangle$.

   c) Calcule la función conjugada de $f = \|\cdot\|$.

3. SUBDIFERENCIAL DE LA NORMA

   Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert. Demostrar que $\partial\|\cdot\|(0) = \mathbb{B}_{\mathbf{X}}$ y que en general se tiene que

   $$\partial\|\cdot\|(x) = \{x^* \in \mathbf{X} \mid \ \|x^*\|_* \leq 1,\ \langle x^*, x\rangle = \|x\|\}, \quad \forall x \in \mathbf{X}.$$

4. INF-CONVOLUCIÓN

   Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert, $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ y $g : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ funciones propias convexas y s.c.i.. Se define la *inf-convolución* de f y g mediante

   $$(f \square g)(x) := \text{ínf}\{f(x_1) + g(x_2) \mid x_1 + x_2 = x\}, \quad \forall x \in \mathbf{X}.$$

   a) Pruebe que $f \square g$ es convexa, con $\mathrm{dom}(f \square g) = \mathrm{dom}(f) + \mathrm{dom}(g)$.

   b) Pruebe que si $\bar{x}_1 \in \mathrm{dom}(f)$ y $\bar{x}_2 \in \mathrm{dom}(g)$ son tales que $(f \square g)(\bar{x}_1 + \bar{x}_2) = f(\bar{x}_1) + g(\bar{x}_2)$, entonces $\partial(f \square g)(\bar{x}_1 + \bar{x}_2) = \partial f(\bar{x}_1) \cap \partial g(\bar{x}_2)$.

   c) (*Efecto Regularizante*) Suponga que $\bar{x}_i$ son los considerados en la parte anterior. Asumiendo que $f \square g$ es subdiferenciable en $\bar{x} = \bar{x}_1 + \bar{x}_2$, muestre que $f \square g$ es Gâteaux-diferenciable en $\bar{x}$ si g lo es en $\bar{x}_2$ con

   $$\nabla(f \square g)(\bar{x}) = \nabla g(\bar{x}_2).$$

   Muestre si además g es Fréchet-diferenciable en $\bar{x}_2$, entonces $f \square g$ también lo es en $\bar{x}$.

   d) Suponga que $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ es un espacio de Hilbert y sea $\mathbf{S} \subset \mathbf{X}$ un conjunto convexo, cerrado, no vacío. Calcular $\partial\,\mathrm{dist}(x, \mathbf{S})$ para $x \notin \mathbf{S}$, donde $x \mapsto \mathrm{dist}(x, \mathbf{S})$ es la función distancia al conjunto $\mathbf{S}$ (ver Ejercicio 6 - Capítulo 3)

5. PROPIEDADES DEL OPERADOR prox

   Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ es un espacio de Hilbert, $\alpha > 0$ y $f\colon \mathbf{X} \mapsto \mathbb{R}\cup\{+\infty\}$ una función propia, convexa y s.c.i. Demuestre que, para todo x e y en $\mathbf{X}$, se tiene

   a) $x - \operatorname{prox}_{\alpha f}(x) \in \alpha\partial f(\operatorname{prox}_{\alpha f}(x))$.

   b) $\|\operatorname{prox}_{\alpha f}(x) - \operatorname{prox}_{\alpha f}(y)\|^2 \leq \|x-y\|^2 - \|(I-\operatorname{prox}_{\alpha f})(x) - (I-\operatorname{prox}_{\alpha f})(y)\|^2$.

   c) $\left\langle \operatorname{prox}_{\alpha f}(x) - \operatorname{prox}_{\alpha f}(y), x-y\right\rangle \geq \|\operatorname{prox}_{\alpha f}(x) - \operatorname{prox}_{\alpha f}(y)\|^2$.

   d) $x \in \arg\min_{\mathbf{X}}(f) \quad \Leftrightarrow \quad x = \operatorname{prox}_{\alpha f}(x)$.

6. EJEMPLOS DE CÁLCULO EXPLÍCITO DEL OPERADOR prox

   a) Sean $X_1, \ldots, \mathbf{X}_n$ espacio de Hilbert y considere $\mathbf{X} = \mathbf{X}_1 \times \cdots \times \mathbf{X}_n$ el espacio producto dotado por el producto interno estándar y $f\colon \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$ definido por

   $$f(x) := \sum_{i=1}^{n} f_i(x_i), \quad \forall x = (x_1, \ldots, x_n) \in \mathbf{X},$$

   donde $f_i : \mathbf{X}_i \to \mathbb{R}\cup\{+\infty\}$ son funciones propias convexas y s.c.i.. Muestre que para todo $\alpha > 0$ se tiene que

   $$\operatorname{prox}_{\alpha f}(x) = (\operatorname{prox}_{\alpha f_1}(x_1), \ldots, \operatorname{prox}_{\alpha f_n}(x_n)),, \quad \forall x = (x_1, \ldots, x_n) \in \mathbf{X}.$$

   Encontrar una expresión explícita para $f(x) = \|x\|_1 = \sum_{i=1}^n |x_i|$, en el caso $\mathbf{X} = \mathbb{R}^n$.

   b) Sea $\mathbf{S} \subset \mathbf{X}$ un conjunto convexo, cerrado, no vacío de un espacio de Hilbert $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ y sea $f = \delta_{\mathbf{S}}$. Muestre que para todo $\alpha > 0$ se tiene $\operatorname{prox}_{\alpha f}(x) = \operatorname{proy}(x, \mathbf{S})$ para todo $x \in \mathbf{X}$.

7. MÉTODO DE EXTRA-GRADIENTES DE KORPELEVICH

   Sean $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert real de dimensión finita y $f : \mathbf{X} \to \mathbb{R}\cup\{+\infty\}$ una función propia convexa y s.c.i. con $\arg\min_{\mathbf{X}}(f) \neq \emptyset$. Consideremos $\sigma \in (0,1)$ y una sucesión $\{\alpha_k\} \subseteq \mathbb{R}$ tal que

   $$\alpha_k > 0, \quad \forall k \in \mathbb{N} \qquad \text{y} \quad \sum_{k=0}^{\infty} \alpha_k = +\infty.$$

   El método de *de extra-gradientes de Korpelevich* consiste en construir recursivamente la secuencia

   $$x_{k+1} = x_k - \alpha_k x_k^*$$

   a partir de un punto inicial $x_0 \in \mathbf{X}$ donde la dirección x_k^* se escoge de forma tal que

   $$x_k^* \in \partial f(y_k) \quad \text{ para algún } y_k \in \mathbf{X} \text{ que satisface } |y_k - x_k + \alpha_k x_k^*| \leq \sigma |y_k - x_k|.$$

   Se propone estudiar la convergencia de este método, para ello proceda como sigue:

   a) Sea $\theta_k = \frac{1}{2}|x_k - \bar{x}|^2$ con $\bar{x} \in \arg\min_{\mathbf{X}}(f)$. Probar la desigualdad

   $$\theta_{k+1} - \theta_k \leq \alpha_k[f(\bar{x}) - \varphi(y_k)] + \frac{1}{2}(\sigma^2 - 1)|y_k - x_k|^2.$$

   b) Deducir que $|y_k - x_k| \to 0$ y concluir que $x_k \to x_\infty$ para algún $x_\infty \in\in \arg\min_{\mathbf{X}}(f)$.

PARTE II

TEORÍA LOCAL DE OPTIMIZACIÓN

Caso general

Resumen. En esta parte del curso nos enfocares en estudiar problemas generales de optimización, no necesariamente convexos. Veremos que la principal diferencia en este caso es que el análisis es esencialmente local y que las condiciones necesarias de optimalidad pueden no ser suficientes. Esta parte del curso se dividirá en dos. En una primera instancia estudiaremos problemas sin restricciones, que será el análogo al capítulo de *Optimización Convexa Diferenciable*. Luego pasaremos a problemas de Programación Matemática donde repasaremos las condiciones de optimalidad de Kuhn-Tucker.

CAPÍTULO 6

Optimización irrestricta

Abstract. En este capítulo estudiaremos problemas de optimización donde se busca minimizar una función diferenciable, no necesariamente convexa. Estudiaremos las condiciones de optimalidad, necesarias y suficientes para que un punto sea un mínimo en un sentido local. Introduciremos además algunos métodos algorítmicos para encontrar mínimos locales de funciones diferenciables.

La optimización convexa entrega una buena intuición sobre lo que es la optimización en general, y de alguna forma puede ser vista como el caso más favorable que uno puede estudiar. A partir de ahora usaremos esa intuición para analizar problemas más generales.

A lo largo de este capítulo trabajaremos básicamente con funciones $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ definidas sobre un espacio de vectorial normado $(\mathbf{X}, \|\cdot\|)$, que para muchos efectos será tomado simplemente como $\mathbb{R}^n$ dotado de la norma Euclideana, que hemos denotado hasta ahora por $|\cdot|$. En general, y de forma similar a lo hecho en el capítulo 4 , asumiremos que $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ es, al menos, Gâteaux diferenciable en $\text{int}(\text{dom}(f))$.

6.1. Mínimos locales

Recordemos que la inf-compacidad y la semi-continuidad inferior son criterios que nos permiten determinar la existencia de mínimos de un problema de optimización del estilo

$$\text{(P)} \qquad \text{Minimizar } f(x) \text{ sobre todos los } x \in \mathbf{X}.$$

En el caso de optimización convexa tenemos que la Regla de Fermat (Teorema 4.4) permite caracterizar los mínimos de una función convexa. Sin embargo, en el caso no convexo esto puede fallar y en general esa regla solo nos entrega información local de la función.

Definición 6.1 (Mínimos locales)**.** *Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función dada. Un punto $\bar{x} \in int(dom(f))$ se dice mínimo local de f si existe $r > 0$ tal que*

$$f(\bar{x}) \leq f(x), \qquad \forall x \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r).$$

Un mínimo local se dice estricto *si la relación anterior es válida con desigualdad estricta.*

Además, $\bar{x} \in \mathbf{X}$ se dirá máximo local (estricto) de f si $\bar{x}$ es un mínimo local (estricto) de $-f$.

En adelante, y para no generar confusión, a los mínimos de una función le agregaremos el adjetivo *global* para distinguirlo de los mínimos locales. Es claro que todo mínimo global del problema (P) es también un mínimo local; de hecho podemos tomar $r > 0$ arbitrario. Sin embargo, como muestra el siguiente ejemplo, mínimos locales no son necesariamente mínimos globales de la función en cuestión, de hecho, su existencia no asegura siquiera que la función sea acotada inferiormente.

Ejemplo 6.1.1. *Consideremos la función sobre $\mathbb{R}$ definida por $f(x) = x^2 - x^4$. No es difícil ver que $\bar{x} = 0$ es un mínimo local de f. Efectivamente, la desigualdad*

$$f(0) = 0 \leq x^2 - x^4$$

es trivial bajo la condición $|x| < 1$ (de hecho es un mínimo local estricto). Luego, $\bar{x} = 0$ es un mínimo local de f pero no es un mínimo global de f, puesto que $f(x) < 0$ para cualquier $|x| > 1$. Más aún, se verifica que $f(x) \to -\infty$ si $|x| \to +\infty$, es decir, f no es acotada inferiormente.

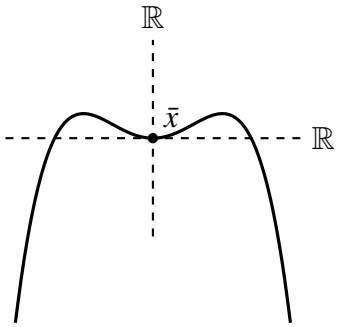


Figura 6.1: Grafo de la función $f(x) = x^2 - x^4$.

La primera gran diferencia que existe entre la optimización convexa y el caso general es que, contrariamente a lo mostrado en el ejemplo anterior, mínimos locales de funciones convexas son también mínimos globales.

Proposición 6.1. *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función convexa dada. Si $\bar{x} \in \mathbf{X}$ es un mínimo local de f entonces $\bar{x} \in \arg\min_{\mathbf{X}}(f)$.*

Demostración. Como $\bar{x}$ es mínimo local, existe $r > 0$ tal que

$$f(\bar{x}) \leq f(x), \quad \forall x \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r).$$

Sea $y \in \mathbf{X}$ y probemos que $f(\bar{x}) \leq y$. Si $y \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r)$ no hay nada que probar, así que supongamos que $\|y - \bar{x}\| > r$ y definamos $z = \bar{x} + \frac{r}{2\|y-\bar{x}\|}(y - \bar{x})$ donde $\frac{r}{2\|y-\bar{x}\|} \in (0,1)$ y, además, $z \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r)$. Luego $f(\bar{x}) \leq f(z)$ y, por convexidad de f, se tiene

$$f(z) \leq f(\bar{x}) + \frac{r}{2\|y - \bar{x}\|}(f(y) - f(\bar{x})),$$

de donde $f(y) - f(\bar{x}) \geq \frac{2\|y-\bar{x}\|}{r}(f(z) - f(\bar{x})) \geq 0$ de donde se deduce el resultado. □

6.2. Condiciones necesarias de optimalidad

La segunda gran diferencia entre la optimización convexa y el caso general se refiere a las condiciones de optimalidad. Recordemos que en el caso convexo diferenciable la Regla de Fermat (Teorema 4.4) dice que un mínimo (global) de $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ está caracterizado como solución de la

ecuación $\nabla f(\bar{x}) = 0$. En el caso general esto es solamente una condición necesaria, pero no suficiente; por ejemplo la función $x \mapsto x^3$ satisface la condición en $\bar{x} = 0$, pero $\bar{x}$ no es un mínimo (global ni local) de la función.

A continuación estudiaremos condiciones necesarias de optimalidad, similares a la Regla de Fermat. Dado que éstas involucran las derivadas de la función objetivo, nos bastará conocer el comportamiento de la función en una vecindad del mínimo en cuestión. Por esta razón las condiciones de optimalidad se puede obtener para mínimos locales y no solamente para mínimos globales.

6.2.1. Condiciones de primer orden

Estudiaremos primero condiciones que involucran la información de primer orden de la función objetivo, es decir, nos bastará conocer la derivada de la función en cuestión.

Teorema 6.1 (Condición necesaria de primer orden)**.** *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función Gâteaux diferenciable en una vecindad de* $\bar{x} \in \mathbf{X}$*. Si* $\bar{x}$ *es un mínimo local de* f *entonces*

(CNPO) $$\nabla f(\bar{x}) = 0.$$

Demostración. Sea $r > 0$ tal que

$$f(\bar{x}) \leq f(x), \quad \forall x \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r)$$

y sea $d \in \mathbf{X} \setminus \{0\}$. Para todo $t < r/\|d\|$ se tiene que $\bar{x} + td \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r)$ y, luego,

(6.1) $$\frac{f(\bar{x}+td) - f(\bar{x})}{t} \geq 0.$$

Tomado límite $t \to 0^+$ se concluye $\langle \nabla f(\bar{x}), d \rangle \geq 0$ para todo $d \in \mathbf{X} \setminus \{0\}$ y el resultado se concluye reemplazando d por $-d$ en el razonamiento anterior. □

Notemos primero que (CNPO), para el caso de espacios de Hilbert y funciones Gâteaux diferenciables se limita simplemente a la condición

$$\nabla f(\bar{x}) = 0.$$

Por otra parte, en la demostración del Teorema 6.1 podríamos cambiar f por $-f$ y obtener la misma conclusión. Esto quiere decir que (CNPO) es ciega con respecto a la operación que se está ejecutando, ya sea minimizar o maximizar. Además, como mencionamos anteriormente en el ejemplo de la función $x \mapsto x^3$, hay puntos que pueden satisfacer (CNPO) y no ser ni mínimos ni máximos de una función. Con el fin de abarcar todo estas clases de puntos introducimos la siguiente definición.

Definición 6.2 (Puntos críticos)**.** *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbertespacio vectorial normado y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función Gâteaux diferenciable en* $int(dom(f))$*. Diremos que un punto* $\bar{x} \in int(dom(f))$ *es un* punto crítico *de* f *si satisface* (CNPO)*, es decir,* $\nabla f(\bar{x}) = 0$*.*

Ejemplo 6.2.1. *Consideremos la función* $f : \mathbb{R} \to \mathbb{R}$ *definida por* $f(x) = \frac{1}{5}x^5 - \frac{1}{3}x^3$*. Esta función tiene tres puntos críticos* $\bar{x}_1 = -1$*,* $\bar{x}_2 = -1$ *y* $\bar{x}_3 = 1$*. Del grafo de la función podemos concluir que,* $\bar{x}_1$ *es un máximo local y* $\bar{x}_3$ *es un mínimo local. El punto* $\bar{x}_2$ *no es ni mínimo ni máximo local.*

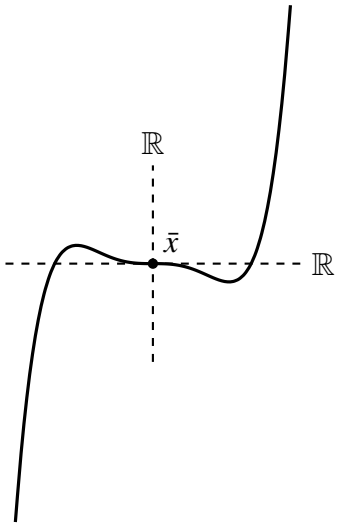


Figura 6.2: Grafo de la función $f(x) = \frac{1}{5}x^5 - \frac{1}{3}x^3$.

6.2.2. Condiciones de segundo orden

Notemos que si la función restringida a una vecindad de un punto crítico fuese convexa entonces la posibilidad que el punto crítico sea un mínimo local aumentan, pues podríamos descartar inmediatamente que ese punto no es un máximo local estricto. Por lo tanto, para poder distinguir y clasificar puntos críticos se requiere más información sobre la función, en particular sobre su curvatura. Veremos ahora un criterio de segundo orden, que simula en cierto grado la convexidad local de una función. Recordemos que una función dos veces Gâteaux diferenciable es convexa si y sólo si $\nabla^2 f(x)$ es un operador bilineal continuo semi-definido positivo.

Teorema 6.2 (Condición necesaria de segundo orden)**.** *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función dos veces Gâteaux diferenciable en una vecindad de* $\bar{x} \in \mathbf{X}$*. Si* $\bar{x}$ *es un mínimo local de* f *entonces* $\nabla f(\bar{x}) = 0$ *y* $\nabla^2 f(\bar{x})$ *es semi-definido positivo, es decir,*

(CNSO) $$\nabla^2 f(\bar{x})(h,h) \geq 0, \quad \forall h \in \mathbf{X}.$$

Demostración. Sea $r > 0$ tal que f es dos veces Gâteaux diferenciable en $\mathbb{B}_{\mathbf{X}}(\bar{x}, r)$ y

$$f(\bar{x}) \leq f(x), \quad \forall x \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r),$$

sea $h \in \mathbf{X} \setminus \{0\}$ (si $h = 0$ no hay nada que probar) y definamos $\phi\colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ dada por

$$\phi(t) = f(\bar{x} + th), \quad \forall t \in \mathbb{R}.$$

Dado que f es dos veces Gâteaux diferenciable en una vecindad de $\bar{x}$, tenemos que ϕ es dos veces derivable en una vecindad de $t = 0$, y por lo tanto

$$f(\bar{x}+th) = \phi(t) = \phi(0) + \phi'(0)t + \phi''(0)\frac{t^2}{2} + o(t^2) = f(\bar{x}) + \langle \nabla f(\bar{x}), th \rangle + \nabla^2 f(\bar{x})(h,h)\frac{t^2}{2} + o(t^2),$$

donde $\lim_{s\to 0} o(s)/s = 0$. Del Teorema 6.1 se deduce $\nabla f(\bar{x}) = 0$ y, luego

$$0 \leq \frac{2(f(\bar{x}+th) - f(\bar{x}))}{t^2} = \nabla^2 f(\bar{x})(h,h) + \frac{o(t^2)}{t^2}.$$

El resultado final se obtiene tomando entonces límite $t \to 0$. □

Es importante destacar que (CNSO) para el caso $\mathbf{X} = \mathbb{R}^n$ y f dos veces Gâteaux diferenciable se traduce en

$$\nabla f(\bar{x}) = 0 \quad \text{y} \quad \nabla^2 f(\bar{x}) \in \mathbb{S}^n_+(\mathbb{R}),$$

donde $\nabla^2 f(\bar{x})$ es la matriz Hessiana de la función f en el punto $\bar{x}$. En otras palabras, para utilizar (CNSO) en este caso es útil conocer los valores propios de la matriz $\nabla^2 f(\bar{x})$; si todos ellos son no negativos, entonces podemos concluir que $\nabla^2 f(\bar{x}) \in \mathbb{S}^n_+(\mathbb{R})$.

Ejemplo 6.2.2. *Retomemos los datos del Ejemplo 6.2.1. En este caso tenemos que* $\nabla^2 f(x) = 4x^3 - 2x$. *Dado que* $\nabla^2 f(-1) = -2$ *podemos inmediatamente descartar el punto* $\bar{x}_1 = -2$ *como mínimo local. Notemos que* $\nabla^2 f(0) = 0$ *por lo que no podemos descartar analíticamente el punto* $\bar{x}_2 = 0$ *como mínimo o máximo local. Además, efectivamente tenemos que* $\nabla^2 f(1) = 2 > 0$ *por lo que el punto* $\bar{x}_3 = 1$ *es candidato a ser mínimo local.*

Ejemplo 6.2.3. *Consideremos la función* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida por* $f(x,y) = x^2(1+y) + y^2/2$. *Esta función tiene tres puntos críticos* $(\bar{x}_1, \bar{y}_1) = (0,0)$ *y* $(\bar{x}_2, \bar{y}_2) = (1,-1)$ *y* $(\bar{x}_3, \bar{y}_3) = (-1,-1)$. *Además, la matriz Hessiana está dada por*

$$\nabla^2 f(x,y) = \begin{pmatrix} 2(1+y) & 2x \\ 2x & 1 \end{pmatrix}$$

De aquí concluimos que $(\bar{x}_1, \bar{y}_1) = (0,0)$ *es candidato a ser mínimo local, pues los valores propios de* $\nabla^2 f(\bar{x}_1, \bar{y}_1)$ *son 1 y 2. Además, (CNSO) nos permite también descarta los puntos* $(\bar{x}_2, \bar{y}_2)$ *y* $(\bar{x}_3, \bar{y}_3)$, *pues la matriz Hessiana en este caso tiene un valor propio positivo y otro negativo (en ambos casos).*

6.3. Condiciones suficientes de optimalidad

Notemos que (CNSO) no logra descartar todos los puntos críticos que no son mínimos locales debido a que ésta es una condición puntual que no puede ser extendida a una vecindad de un punto crítico $\bar{x}$. Es decir, la condición que el operador $\nabla^2 f(\bar{x})$ sea semi-definido positivo no implica necesariamente que $\nabla^2 f(x)$ sea también semi-definido positivo para todo $x \in \mathbf{X}$ que pertenezca a una vecindad de $\bar{x}$. Para obtener una condición de este estilo necesitamos hacer más fuerte (CNSO). Como consecuencia obtenemos un resultado más fuerte, que logra no sólo clasificar a un punto crítico como mínimo local, si no que además como mínimo local estricto.

Teorema 6.3 (Condición suficiente de segundo orden)**.** *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbert y* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función dos veces Fréchet diferenciable en una vecindad de* $\bar{x} \in \mathbf{X}$. *Supongamos que* $\bar{x}$ *es un punto crítico de* f *y que existe* $\alpha > 0$ *tal que*

$$\nabla^2 f(\bar{x})(h,h) \geq \alpha \|h\|^2, \quad \forall h \in \mathbf{X}. \tag{CSSO}$$

Entonces $\bar{x}$ *es un mínimo local estricto de* f.

Demostración. Sea $r > 0$ tal que f es dos veces Fréchet-diferenciable en $\mathbb{B}_{\mathbf{X}}(\bar{x}, r)$. Primero probaremos que, para $h \in \mathbb{B}_{\mathbf{X}}(0, r)$

$$f(\bar{x}+h) - f(\bar{x}) - \langle \nabla f(\bar{x}), h \rangle - \frac{1}{2}\nabla^2 f(\bar{x})(h,h) = o(\|h\|^2).$$

En efecto, llamando

$$\varphi(h) := f(\bar{x}+h) - f(\bar{x}) - \langle \nabla f(\bar{x}), h \rangle - \frac{1}{2}\nabla^2 f(\bar{x})(h,h),$$

por simetría de $\nabla^2 f(\bar{x})(\cdot,\cdot)$ se tiene

$$\langle \nabla \varphi(h), k \rangle = \langle \nabla f(\bar{x}+h) - \nabla f(\bar{x}), k \rangle - \nabla^2 f(\bar{x})(h,k)$$

y la Fréchet diferenciablilidad de segundo orden implica

$$\lim_{h\to 0} \frac{\|\nabla \varphi(h)\|}{\|h\|} = 0.$$

De ese modo, como $\varphi(0) = 0$ se tiene del Teorema del Valor Medio aplicado a $t \mapsto \varphi(th)$ que existe $\lambda \in (0,1)$ tal que

$$|\varphi(h)| = |\varphi(h) - \varphi(0)| = |\langle \nabla \varphi(\lambda h), h \rangle| \leq \|\nabla \varphi(\lambda h)\| \|h\|, \tag{6.2}$$

de donde

$$\frac{|\varphi(h)|}{\|h\|^2} \leq \frac{\|\nabla \varphi(\lambda h)\|}{\|h\|} \leq \frac{\|\nabla \varphi(\lambda h)\|_*}{\|\lambda h\|}$$

y el resultado se obtiene tomando $h \to 0$.

Por lo tanto, usando este resultado y que $\nabla f(\bar{x}) = 0$ por ser punto crítico, se obtiene

$$f(\bar{x}+h) - f(\bar{x}) = \frac{1}{2}\nabla^2 f(\bar{x})(h,h) + o(\|h\|^2) \geq \frac{\alpha}{2}\|h\|^2 + o(\|h\|^2)$$

y tomando $r > 0$ tal que $o(\|h\|^2)/\|h\|^2 \leq \alpha/4$ para todo $h \in \mathbb{B}_{\mathbf{X}}(0,r) \setminus \{0\}$ se deduce

$$\frac{f(\bar{x}+h) - f(\bar{x})}{\|h\|^2} \geq \alpha/4 > 0, \quad \forall h \in \mathbb{B}_{\mathbf{X}}(0,r) \setminus \{0\}.$$

□

Ejemplo 6.3.1. *Retomando los datos del Ejemplo 6.2.3, tenemos que de todos los puntos críticos de la función, solamente el punto $(\bar{x}_1, \bar{y}_1) = (0,0)$ es candidato a ser mínimo local. Como ya vimos, $\nabla^2 f(\bar{x}_1, \bar{y}_1)$ tiene valores propios positivos, es decir, es una matriz definida positiva. Por lo tanto, usando (CSSO) podemos concluir que $(\bar{x}_1, \bar{y}_1)$ es un mínimo local estricto de la función en cuestión.*

6.4. Métodos de Direcciones de Descenso

Ahora estudiaremos algunos métodos iterativos para encontrar mínimos locales de funciones Gâteaux diferenciables. Por simplicidad de la exposición nos restringiremos al caso $\mathbf{X} = \mathbb{R}^n$, donde la norma será la norma Euclideana. La idea principal de los métodos que presentaremos es que nos permitirán construir sucesiones $\{x_k\}$ en $\mathbb{R}^n$ tal que $\nabla f(x_k) \to 0$ cuando $k \to +\infty$.

La forma general de los métodos que estudiaremos se basa en una iteración del tipo

$$x_{k+1} = x_k + \alpha_k d_k, \quad \forall k \in \mathbb{N} \tag{6.3}$$

que parte desde $x_0 \in \mathbb{R}^n$, donde $\alpha_k > 0$ y $d_k \in \mathbb{R}^n$ son tales que nos aseguran que $f(x_{k+1}) < f(x_k)$.

6.4.1. Direcciones de descenso

La principal característica de los métodos que estudiaremos es que la elección de las direcciones d_k se hace de forma tal que asegura la existencia de, al menos, un $\alpha_k > 0$ tal que $f(x_k + \alpha_k d_k) < f(x_k)$. Por esta razón la siguiente definición nos será de utilidad.

Definición 6.3. *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función Gâteaux diferenciable en $x \in int(dom(f))$. Diremos que $d \in \mathbb{R}^n$ es una dirección de descenso de f en x si*

$$\nabla f(x)^\top d < 0.$$

Definimos también el ángulo de descenso *de f en el punto x en la dirección d, denotado $\theta_f(x,d)$, via la ecuación*

$$\cos\left(\theta_f(x,d)\right) = \frac{-\nabla f(x)^\top d}{|\nabla f(x)||d|} \in (0,1].$$

Notemos que si d_k es una dirección de descenso para f en x_k, entonces si $f(x_k) + \alpha \nabla f(x_k)^\top d_k$ es un buena aproximación de $f(x_k + \alpha d_k)$ para $\alpha \simeq 0$ (Taylor de primer orden), entonces la existencia de $\alpha_k > 0$ tal que $f(x_k + \alpha_k d_k) < f(x_k)$ queda asegurada.

Observación 6.1. *Hasta ahora hemos visto, para el caso convexo y bajo condiciones apropiadas, tres ejemplos donde d_k es una dirección de descenso (ver Sección 4.5):*

- *Método del Gradiente: $d_k = -\nabla f(x_k)$ con $\cos\left(\theta_f(x_k, d_k)\right) = 1$.*
- *Método del Gradiente conjugado: $d_k = -\nabla f(x_k) + \beta_k d_{k-1}$*

  En este caso es esencial el hecho que $\alpha_k > 0$ se escoge usando la regla de búsqueda lineal exacta, es decir, α_k minimiza la función $\alpha \mapsto f(x_k + \alpha d_k)$, pues esto implica a su vez que $\nabla f(x_k)^\top d_{k-1} = 0$ para todo $k \in \mathbb{N} \setminus \{0\}$.
- *Método de Newton-Raphson: $d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$ con $\cos\left(\theta_f(x_k, d_k)\right) \geq \frac{1}{\kappa\left(\nabla^2 f(x_k)\right)}$.*

  Para este caso, es fundamental que $\nabla^2 f(x_k)$ sea definida positiva.

Usando la definición anterior, el algoritmo general de métodos de descenso se escribe como

---

ALGORITMO DE MÉTODO DE DIRECCIONES DE DESCENSO

Supongamos que conocemos $x_k \in \mathbb{R}^n$

1. **Criterio de parada:** si $\nabla f(x_k) \simeq 0$, parar.
2. **Dirección de descenso:** escoger una dirección de descenso $d_k \in \mathbb{R}^n$.
3. **Búsqueda lineal:** determinar un paso $\alpha_k > 0$ de forma tal que f *decrezca suficientemente* en la dirección d_k.
4. **Actualización:** $x_{k+1} = x_k + \alpha_k d_k$.

Otra dirección de descenso que vale la pena mencionar, y que estudiaremos en profundidad más adelante, es la dirección de descenso del Método *Quasi-Newton*, la cuál se inspira en el método de Newton-Raphson. La idea principal es tomar la dirección de descenso de la forma

$$d_k = -B_k^{-1} \nabla f(x_k), \quad \forall k \in \mathbb{N}$$

donde $B_k \in \mathbb{S}^n_{++}(\mathbb{R})$ es una matriz que aproxima a $\nabla^2 f(x_k)$ en algún sentido. Notemos además que

$$\cos\left(\theta_f(x_k, d_k)\right) \geq \frac{1}{\kappa(B_k)}, \quad \forall k \in \mathbb{N}.$$

6.4.2. Reglas de Búsqueda Lineal inexactas

Ahora nos enfocaremos en formas de determinar el paso $\alpha_k > 0$ para dar sentido a la frase *que f decrezca suficientemente en la dirección d_k*. Veremos también que estas reglas nos servirán para estudiar la convergencia del algoritmo.

La forma más natural de determinar un paso $\alpha_k > 0$ es simplemente tomar $\alpha_k = \bar{\alpha}$, donde $\bar{\alpha}$ minimiza la función $\alpha \mapsto f(x_k + \alpha d_k)$. Esto se conoce como la *regla de búsqueda lineal exacta*. Hemos visto que para problemas cuadráticos estrictamente convexo se puede encontrar una fórmula explícita para α_k. Desafortunadamente, en el caso no lineal general, calcular α_k puede ser muy difícil y normalmente no se obtienen fórmulas explícitas; una de las dificultades es que la Regla de Fermat es una ecuación no lineal difícil de resolver. Por esta razón es mejor enfocarse en reglas de búsqueda lineal inexacta, es decir, donde α_k no es el óptimo de $\alpha \mapsto f(x_k + \alpha d_k)$, pero satisface dos condiciones esenciales: (i) hace decrecer la función $\alpha \mapsto f(x_k + \alpha d_k)$ de forma razonable, y (ii) no requiere demasiado tiempo ni esfuerzo para ser calculado.

La idea detrás de estas reglas de búsqueda es intentar con un serie de candidatos hasta que uno satisfaga una condición que asegure un decrecimiento sustancial de la función en la dirección d_k.

Regla de Armijo

La primera regla de búsqueda lineal inexacta que estudiaremos, llamada *regla de Armijo*, consiste en pedir que el decrecimiento sea proporcional a un cierto $\omega_1 \in (0,1)$. Esto se traduce en que la función decrece de forma lineal en la dirección d_k. Dicho de otra forma, la *condición de Armijo* pide que $\alpha_k > 0$ se escoja de forma tal que

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \omega_1 \alpha_k \nabla f(x_k)^\top d_k. \tag{6.4}$$

Notemos que ω_1 está fijo en la condición de Armijo (no cambia con k) y a priori no hay mayor restricción sobre él. Sin embargo, en la práctica, y con el fin que (6.4) sea más fácil de verificar, se toma ω_1 pequeño (típicamente $\omega_1 \simeq 10^{-4}$).

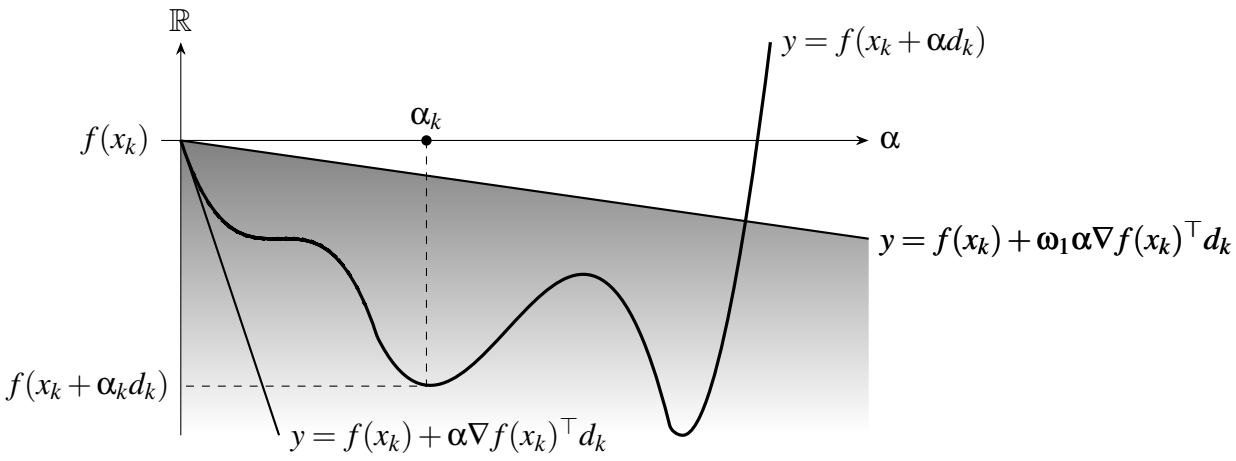


Figura 6.3: Regla de Armijo.

Para encontrar un paso que satisfaga la condición de Armijo se procede en general usando una técnica llamada *backtracking* y que está determinada por el siguiente algoritmo:

REGLA DE ARMIJO (BACKTRACKING)

1. Tomar $\alpha > 0$, $\tau \in \left(0, \frac{1}{2}\right)$.
2. Si α satisface la regla de Armijo, fijar $\alpha_k = \alpha$ y parar.
3. Escoger $\beta \in [\tau\alpha, (1-\tau)\alpha]$.
4. Actualizar $\alpha = \beta$ y volver al paso 2.

Normalmente, τ es pequeño (en general $10^{-2} \leq \tau \leq 10^{-1}$) y esta elección de pasos se asocia frecuentemente a direcciones de descenso de Newton-Raphson, pues en este caso se espera tener convergencia con $\alpha_k \simeq 1$.

Regla de Goldstein

Notemos que la elección del paso con la regla de Armijo no provee una cota inferior para el paso, y no hay en principio mayor inconveniente en escoger α_k muy pequeño. El problema es que esto puede llevar a que el algoritmo converja a un punto que no es necesariamente un punto crítico de la función. En efecto, si el paso se escoge de forma tal que para algún $\varepsilon > 0$ se cumple

$$0 < \alpha_k \leq \frac{\varepsilon}{2^{k+1}|d_k|}, \quad \forall k \in \mathbb{N}$$

Tendremos que la sucesion $\{x_k\}$ generada por (6.3) es de Cauchy y por lo tanto converge a algún $\bar{x} \in \mathbb{R}^n$. En efecto para todo $l \in \mathbb{N}$ tenemos

$$|x_{k+l} - x_k| = \left| \sum_{i=k}^{k+l-1} \alpha_i d_i \right| \leq \sum_{i=k}^{\infty} \frac{\varepsilon}{2^{i+1}} \to 0 \quad \text{si} \quad k \to +\infty.$$

Esto a su vez implica que

$$|\bar{x} - x_0| \leq \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon.$$

Por lo tanto, si no hay puntos crítico de f cerca de x_0, entonces $\bar{x}$ no puede ser un punto crítico -ni mínimo local- de f. Es decir, en este caso, el Método de Direcciones de Descenso podría no converger en el sentido que $\nabla f(x_k) \to \nabla f(\bar{x}) \neq 0$ cuando $k \to +\infty$.

Para evitar este tipo de problemas, se introduce una nueva regla, llamada *regla de Goldstein* y cuyo objetivo es evitar que α_k se escoja muy pequeño. Dicho de otra forma, la *condición de Goldstein* pide que $\alpha_k > 0$ satisfaga

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \omega_1 \alpha_k \nabla f(x_k)^\top d_k. \tag{6.5a}$$

$$f(x_k + \alpha_k d_k) \geq f(x_k) + (1-\omega_1)\alpha_k \nabla f(x_k)^\top d_k. \tag{6.5b}$$

Notar que (6.5a) no es otra cosa que la condición de Armijo (6.4).

El siguiente resultado muestra que siempre es posible escoger un paso según la regla de Goldstein (y en consecuencia según la regla de Armijo).

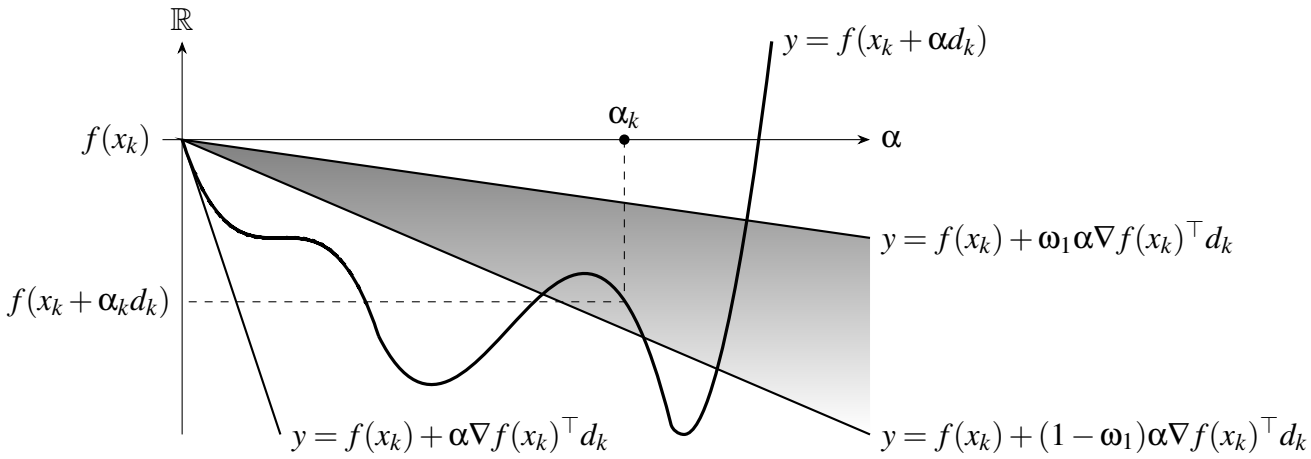


Figura 6.4: Regla de Goldstein.

Proposición 6.2. *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función inferiormente acotada tal que $dom(f)$ es un abierto de $\mathbb{R}^n$. Supongamos además que f es continua y Gâteaux diferenciable en $dom(f)$. Sea $k \in \mathbb{N}$ y x_k una instancia del Método de Direcciones de Descenso (6.3) con d_k siendo una dirección de descenso. Entonces, para todo $\omega_1 \in (0, 1/2)$ existe $\alpha_k > 0$ que verifica la Regla de Goldstein (6.5).*

Demostración. En efecto, dado $x_k \in \text{dom}(f)$, d_k dirección de descenso, por diferenciabilidad de f se tiene

$$\lim_{\alpha \to 0^+} \frac{f(x_k + \alpha d_k) - f(x_k)}{\alpha} = \nabla f(x_k)^\top d_k < (1-\omega_1)\nabla f(x_k)^\top d_k < \omega_1 \nabla f(x_k)^\top d_k.$$

Además, como $\nabla f(x_k)^\top d_k < 0$ debido a que d_k es dirección de descenso,

$$\lim_{\alpha \to \infty} f(x_k + \alpha d_k) \geq \inf_{\mathbb{R}^n}(f) > -\infty = \lim_{\alpha \to \infty} f(x_k) + \alpha\omega_1 \nabla f(x_k)^\top d_k = \lim_{\alpha \to \infty} f(x_k) + \alpha(1-\omega_1)\nabla f(x_k)^\top d_k.$$

Por lo tanto, por continuidad de las funciones $\alpha \mapsto f(x_k + \alpha d_k)$, $\alpha \mapsto f(x_k) + \alpha\omega_1 \nabla f(x_k)^\top d_k$ y $\alpha \mapsto f(x_k) + \alpha(1-\omega_1)\nabla f(x_k)^\top d_k$, se deduce del teorema del valor intermedio que existen $\alpha_2 < \alpha_1$ tales que

$$\alpha_1 = \text{ínf}\{\alpha > 0 \mid f(x_k + \alpha d_k) = f(x_k) + \alpha\omega_1 \nabla f(x_k)^\top d_k\}$$
$$\alpha_2 = \sup\{\alpha \in (0, \alpha_1) \mid f(x_k + \alpha d_k) = f(x_k) + \alpha(1-\omega_1)\nabla f(x_k)^\top d_k\}$$

y por lo tanto las condiciones de Goldstein (6.5) se cumplen para todo $\alpha_2 \leq \alpha_k \leq \alpha_1$.

□

Regla de Wolfe

Otra forma de evitar el problema de convergencia a un punto que no es un punto crítico es introducir una regla que considere información sobre la curvatura de la función. La *condición de Wolfe* pide que $\alpha_k > 0$ satisfaga, para algún $\omega_1 \in (0, 1)$ y $\omega_2 \in (\omega_1, 1)$, las siguientes condiciones

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \omega_1 \alpha_k \nabla f(x_k)^\top d_k. \tag{6.6a}$$

$$\nabla f(x_k + \alpha_k d_k)^\top d_k \geq \omega_2 \nabla f(x_k)^\top d_k. \tag{6.6b}$$

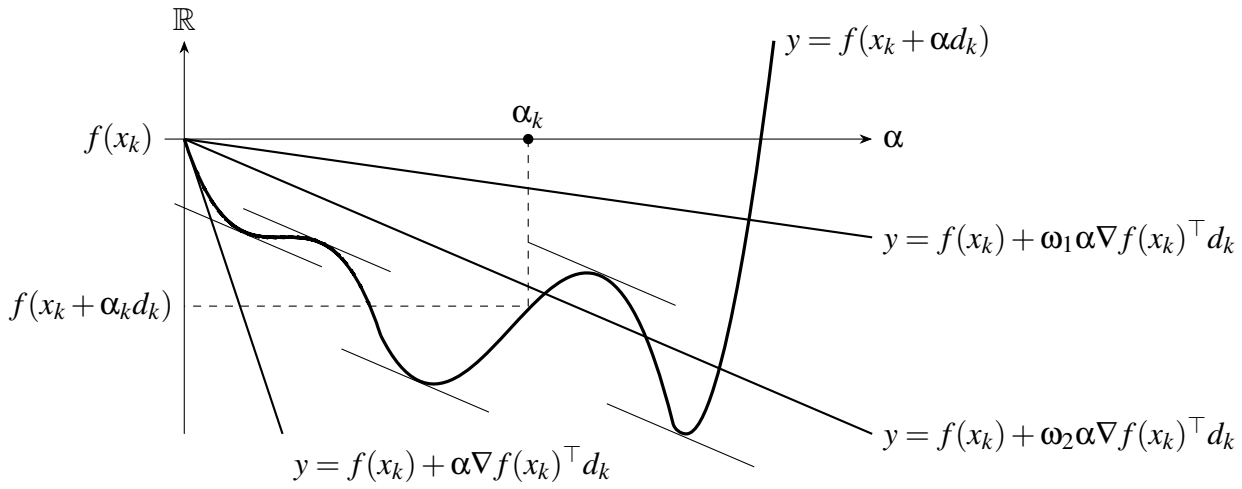


Figura 6.5: Regla de Wolfe.

Notar que, al igual que con la regla de Goldstein, (6.6a) es la condición de Armijo (6.4). Más aún, dado que $\nabla f(x_k + \alpha d_k)^\top d_k$ es la pendiente de la función $\alpha \mapsto f(x_k + \alpha d_k)$ en el punto α, la condición (6.6b) dice que la pendiente $\alpha \mapsto f(x_k + \alpha d_k)$ en α_k debe ser mayor que una proporción ω_2 de la pendiente en $\alpha = 0$, y en consecuencia α_k estará lo suficientemente alejado de $\alpha = 0$ para evitar una falsa convergencia. Notemos además que ω_2, al igual que ω_1 en la condición de Armijo, está fijo (no cambia con k). En la práctica, y con el fin que (6.6b) sea más fácil de verificar, se toma ω_2 cercano a 1 (típicamente $\omega_2 \simeq 0{,}99$). Esta regla, debibo a su relación con la curvatura, se asocia frecuentemente con direcciones de descenso del Método Quasi-Newton.

Veamos ahora que la regla de Wolfe está bien definida.

Proposición 6.3. *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función inferiormente acotada tal que $dom(f)$ es un abierto de $\mathbb{R}^n$. Supongamos que f es continua y Gâteaux diferenciable en $dom(f)$. Sea $k \in \mathbb{N}$ y x_k una instancia del Método de Direcciones de Descenso (6.3) con d_k siendo una dirección de descenso. Entonces, para todo $0 < \omega_1 < \omega_2 < 1$ existe $\alpha_k > 0$ que satisface la condición de Wolfe (6.6).*

Demostración. Consideremos

$$\alpha_1 = \text{ínf}\{\alpha > 0 : f(x_k + \alpha d_k) = f(x_k) + \alpha\omega_1 \nabla f(x_k)^\top d_k\},$$

cuya existencia está garantizada por la demostración de la proposición anterior (Proposición 6.2). Notemos que la primera condición de Wolfe (6.6a) se satisface para todo $\alpha_k \leq \alpha_1$. Por otra parte, por Teorema del Valor Medio, se tiene que existe $\alpha_2 \in (0, \alpha_1)$ tal que

$$\omega_2 \nabla f(x_k)^\top d_k < \omega_1 \nabla f(x_k)^\top d_k = \frac{f(x_k + \alpha_1 d_k) - f(x_k)}{\alpha_1} = \nabla f(x_k + \alpha_2 d_k)^\top d_k \tag{6.7}$$

y por continuidad hay un intervalo alrededor de α_2 donde las condiciones se siguen satisfaciendo. □

Ahora presentaremos un algoritmo (Fletcher-Lemaréchal) que permite encontrar un paso $\alpha_k > 0$ que satisface la condición de Wolfe. Este algoritmo usa igualmente la técnica *backtracking* y se caracteriza por encontrar un paso acorde a la regla de Wolfe en una cantidad finita de iteraciones.

		REGLA DE WOLFE (ALGORITMO DE FLETCHER-LEMARÉCHAL)
1.		Tomar $\alpha > 0$, $\underline{\alpha} = 0$, $\overline{\alpha} = +\infty$, $\tau_i \in \left(0, \frac{1}{2}\right)$ y $\tau_e > 1$.
2.		Si α no satisface (6.6a):
	2.1	Actualizar $\overline{\alpha} = \alpha$
	2.2	Escoger $\beta \in [(1-\tau_i)\underline{\alpha} + \tau_i\overline{\alpha}, \tau_i\underline{\alpha} + (1-\tau_i)\overline{\alpha}]$.
	2.3	Actualizar $\alpha = \beta$
3.		Si α satisface (6.6a):
	3.1	Si α satisface (6.6b), fijar $\alpha_k = \alpha$ y parar.
	3.2	Actualizar $\underline{\alpha} = \alpha$
	3.3	Si $\overline{\alpha} = +\infty$, escoger $\beta \in [\tau_e\underline{\alpha}, +\infty)$.
	3.4	Si $\overline{\alpha} < +\infty$, escoger $\beta \in [(1-\tau_i)\underline{\alpha} + \tau_i\overline{\alpha}, \tau_i\underline{\alpha} + (1-\tau_i)\overline{\alpha}]$.
	3.5	Actualizar $\alpha = \beta$.
4		Volver al paso 2.

Estudiemos ahora la convergencia de este algoritmo.

Proposición 6.4. *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función inferiormente acotada tal que $dom(f)$ es un abierto de $\mathbb{R}^n$. Supongamos además que f es continua y Gâteaux diferenciable en $dom(f)$. Sea $k \in \mathbb{N}$ y x_k una instancia del Método de Direcciones de Descenso (6.3) con d_k siendo una dirección de descenso. Entonces, para todo $0 < \omega_1 < \omega_2 < 1$ el algoritmo de Fletcher-Lemaréchal encuentra un paso $\alpha_k > 0$ que satisface la condición de Wolfe (6.6) en una cantidad finita de pasos.*

6.4.3. Convergencia del Método de Direcciones de Descenso

En esta parte del curso estudiaremos la convergencia del Método de Direcciones de Descenso bajo condiciones bastante generales. Nos enfocaremos en el caso que el paso se escoge usando la regla de Wolfe. Sin embargo cabe destacar que un resultado similar se puede obtener para la regla de Goldstein y Armijo (ésta última con paso acotado uniformemente sobre cero).

Teorema 6.4 (Condición de Zoutendijk)**.** *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función inferiormente acotada continua y Gâteaux diferenciable en $dom(f)$ (abierto de $\mathbb{R}^n$). Supongamos existe un abierto $A \subseteq \mathbb{R}^n$ que contiene al conjunto de subniveles $\Gamma_{f(x_0)}(f)$ para algún $x_0 \in \mathbb{R}^n$, y supongamos que ∇f es L-Lipschitz continua en A. Sea $\{x_k\}$ la sucesión generada por el Método de Direcciones de Descenso (6.3) con d_k siendo una dirección de descenso y α_k dado por la regla de Wolfe (6.6) para $0 < \omega_1 < \omega_2 < 1$. Entonces se tiene*

$$\sum_{k=0}^{\infty} \cos^2(\theta_k)|\nabla f(x_k)|^2 < +\infty,$$

donde $\theta_k = \theta_f(x_k, d_k)$ es el ángulo de descenso de f en el punto x_k en la dirección d_k.

Demostración. Sea $k \in \mathbb{N}$. De la segunda condición de Wolfe, de $x_{k+1} = x_k + \alpha_k d_k$ y del hecho que ∇f es L-Lipschitz se deduce

$$(\omega_2 - 1)\nabla f(x_k)^\top d_k \leq (\nabla f(x_{k+1}) - \nabla f(x_k))^\top d_k \leq L\alpha_k \|d_k\|^2,$$

de donde

$$\alpha_k \geq \frac{(\omega_2 - 1)}{L\|d_k\|^2} \nabla f(x_k)^\top d_k.$$

Ocupando esta desigualdad en la primera condición de Wolfe y usando la definición de θ_k, se deduce

$$f(x_{k+1}) - f(x_k) \leq \frac{\omega_1(\omega_2 - 1)}{L} \left(\frac{\nabla f(x_k)^\top d_k}{\|d_k\|} \right)^2 = -\frac{\omega_1(1-\omega_2)}{L} \cos^2(\theta_k) \|\nabla f(x_k)\|^2,$$

y por lo tanto $\{f(x_k)\}$ es una sucesión real decreciente y acotada inferiormente, y en consecuencia converge. Sumando sobre k se deduce

$$\frac{\omega_1(1-\omega_2)}{L} \sum_{k=0}^{N-1} \cos^2(\theta_k) \|\nabla f(x_k)\|^2 \leq f(x_0) - f(x_N).$$

como el lado derecho converge, la serie es convergente y el resultado se concluye. □

Una consecuencia importante de la condición de Zoutendijk es que si el ángulo de descenso θ_k de f en el punto x_k en la dirección d_k está acotado uniformemente sobre cero, entonces el Método de Direcciones de Descenso converge en el sentido que $\nabla f(x_k) \to 0$.

6.4.4. Método de Newton-Raphson y Quasi-Newton

En adelante estudiaremos en detalle el método Quasi-Newton, en particular en esta parte nos enfocaremos la tasa de convergencia. Luego mostraremos unos métodos para construir las direcciones de descenso (obtener las matrices B_k). Recordemos que la dirección de descenso de *Quasi-Newton* tiene la forma

$$d_k = -B_k^{-1} \nabla f(x_k), \quad \forall k \in \mathbb{N}$$

donde $B_k \in \mathbb{S}^n_{++}(\mathbb{R})$ es una matriz que aproxima a $\nabla^2 f(x_k)$ en algún sentido. Recordemos también que

$$\cos(\theta_k) \geq \frac{1}{\kappa(B_k)}, \quad \forall k \in \mathbb{N}.$$

Por lo tanto, si $\kappa(B_k)$ se mantiene uniformemente acotado superiormente (lo que se traduce en que la sucesión $\{\lambda_{\text{mín}}(B_k)\}$ es uniformemente positiva), entonces la Condición de Zoutendijk, asegura que el método converge. Es claro también, que el Método de Newton-Raphson es una instancia particular del Método Quasi-Newton (basta tomar $B_k = \nabla^2 f(x_k)$), y por lo tanto los resultados que presentaremos a continuación son también válidos para el Método de Newton-Raphson.

Tasa de Convergencia del Método de Newton-Raphson

Recordemos que, en el caso convexo, el Método de Newton-Raphson converge de forma cuadrática (ver Teorema 4.10) cuando la condición inicial está lo suficientemente cerca del mínimo. En este caso, la importancia de la convexidad está en que todo punto crítico es un mínimo global de la función. Si la hipótesis de convexidad de levanta, entonces, dado que la convergencia es sólo local, la convergencia cuadrática sigue siendo cierta, pero el límite es un mínimo local estricto, no necesariamente global. Ahora presentaremos la adaptación al caso no convexo del Teorema 4.10.

Teorema 6.5. *Sea $f\colon \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ una función propia y dos veces Gâteaux diferenciable en $dom(f)$, el cual asumimos ser un abierto de $\mathbb{R}^n$. Supongamos que existe $\bar{x}\in\mathbb{R}^n$ tal que $\nabla f(\bar{x})=0$ y $\nabla^2 f(\bar{x})\in\mathbb{S}^n_{++}(\mathbb{R})$, y que además $\nabla^2 f$ es localmente Lipschitz continua en torno a $\bar{x}$. Entonces, existe $\rho>0$ para el cual se tiene que si $x_0\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$, la secuencia $\{x_k\}$ generada por*

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1}\nabla f(x_k), \qquad \forall k\in\mathbb{N} \tag{6.8}$$

está bien definida, converge a $\bar{x}$ y satisface

$$\lim_{k\to\infty}\frac{|x_{k+1}-\bar{x}|}{|x_k-\bar{x}|} = \lim_{k\to\infty}\frac{|\nabla f(x_{k+1})|}{|\nabla f(x_k)|} = 0, \quad \limsup_{k\to\infty}\frac{|x_{k+1}-\bar{x}|}{|x_k-\bar{x}|^2}<\infty, \quad y \quad \limsup_{k\to\infty}\frac{|\nabla f(x_{k+1})|}{|\nabla f(x_k)|^2}<\infty.$$

Demostración. La primera parte de la demostración sigue el mismo razonamiento que la demostración del Teorema 4.10 y lo único nuevo por probar son los límites con los gradientes. Sin embargo, por claridad de la exposición mostraremos todos los pasos.

Recordemos que, para $x\in\mathrm{dom}(f)$ habíamos denotado por λ_x al menor valor propio de $\nabla^2 f(x)$. Como $\nabla^2 f(\bar{x})\in\mathbb{S}^n_{++}(\mathbb{R})$, de la Proposición 4.2 se tiene

$$y^\top\nabla^2 f(\bar{x})y \geq \lambda_{\bar{x}}|y|^2, \quad \forall y\in\mathbb{R}^n,$$

donde $\lambda_{\bar{x}}>0$. Para todo $x\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},r)$ e $y\in\mathbb{R}^n$, usando la propiedad Lipschitz de $\nabla^2 f$ se tiene

$$\begin{aligned} y^\top\nabla^2 f(x)y &= y^\top\nabla^2 f(\bar{x})y + y^\top(\nabla^2 f(x)-\nabla^2 f(\bar{x}))y \\ &\geq \lambda_{\bar{x}}|y|^2 - \|\nabla^2 f(x)-\nabla^2 f(\bar{x})\||y|^2 \\ &\geq (\lambda_{\bar{x}} - L|x-\bar{x}|)|y|^2. \end{aligned}$$

Luego, definiendo $\rho = \min\left\{r, \frac{\lambda_{\bar{x}}}{2L}\right\}>0$ tenemos

$$\nabla^2 f(x)\in\mathbb{S}^n_{++}(\mathbb{R}) \text{ con } \lambda_x \geq \frac{\lambda_{\bar{x}}}{2} > 0, \quad x\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho).$$

De ese modo, para todo $x\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$, existen matrices P_x y D_x tales que $\nabla^2 f(x) = P_x D_x P_x^\top$ con $P_x^{-1}=P_x^\top$, de modo que $\nabla^2 f(x)^{-1} = P_x D_x^{-1} P_x^\top$ y

$$\|\nabla^2 f(x)^{-1}\| = \frac{1}{\lambda_x} \leq \frac{2}{\lambda_{\bar{x}}}$$

Supongamos que $x_k\in B_{\mathbb{R}^n}(\bar{x},\rho)$ para algún $k\in\mathbb{N}$. Para simplificar la notación, notemos $g_k=\nabla f(x_k)$ y $H_k=\nabla^2 f(x_k)$. De (6.8) se deduce que si $x_k=\bar{x}$ entonces $x_{k+1}=\bar{x}$, por lo que suponemos que $x_k\neq\bar{x}$. Como $\bar{x}$ es un punto crítico de f, es decir, $\nabla f(\bar{x})=0$, usando la propiedad de Lipschitz continuidad de $\nabla^2 f$ y la relación

$$g_k = \nabla f(x_k) - \nabla f(\bar{x}) = \int_0^1 \nabla^2 f(\bar{x}+t(x_k-\bar{x}))(x_k-\bar{x})dt,$$

tenemos que

$$\begin{aligned}
|x_{k+1}-\bar{x}| &= |x_k-\bar{x}-H_k^{-1}g_k| \\
&= |H_k^{-1}\left(H_k(x_k-\bar{x})-g_k\right)| \\
&= \left|H_k^{-1}\left(\int_0^1 [H_k-\nabla^2 f(\bar{x}+t(x_k-\bar{x}))](x_k-\bar{x})dt\right)\right| \\
&\leq \frac{2}{\lambda_{\bar{x}}}|x_k-\bar{x}|\int_0^1 \|H_k-\nabla^2 f(\bar{x}+t(x_k-\bar{x}))\|dt \\
&\leq \frac{2L}{\lambda_{\bar{x}}}|x_k-\bar{x}|^2\int_0^1 (1-t)dt \\
&= \frac{L}{\lambda_{\bar{x}}}|x_k-\bar{x}|^2 \leq \frac{1}{2}|x_k-\bar{x}|,
\end{aligned}$$

En particular, se tiene que $x_{k+1}\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$. Además, usando inducción vemos que la sucesión $\{x_k\}$ está contenida en $\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$ si $x_0\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$ y

$$|x_{k+1}-\bar{x}| \leq \frac{1}{2^{k+1}}|x_0-\bar{x}|, \quad \forall k\in\mathbb{N}.$$

De aquí se concluye que $x_k\to\bar{x}$, y que también tenemos

$$\frac{|x_{k+1}-\bar{x}|}{|x_k-\bar{x}|} \leq \frac{L}{\lambda_{\bar{x}}}|x_k-\bar{x}| \quad \text{y} \quad \frac{|x_{k+1}-\bar{x}|}{|x_k-\bar{x}|^2} \leq \frac{L}{\lambda_{\bar{x}}}.$$

Por otra parte, dado que $H_k(x_{k+1}-x_k)+g_k=0$ para todo $k\in\mathbb{N}$, tenemos que

$$|g_{k+1}| = |g_{k+1}-g_k-H_k(x_{k+1}-x_k)| = \left|\int_0^1 \nabla^2 f(x_k+t(x_{k+1}-x_k))(x_{k+1}-x_k)dt - H_k(x_{k+1}-x_k)\right|.$$

Sigue que,

$$|g_{k+1}| \leq |x_{k+1}-x_k|\int_0^1 \left\|\nabla^2 f(x_k+t(x_{k+1}-x_k))-H_k\right\|dt \leq \frac{L}{2}|x_{k+1}-x_k|^2 \leq \frac{L}{2}\|H_k^{-1}\|^2|g_k|^2.$$

Esto a su vez implica que

$$|g_{k+1}| \leq \frac{4L}{\lambda_{\bar{x}}^2}|g_k|^2, \quad \forall k\in\mathbb{N}.$$

Por lo tanto, usando los mismos argumentos que más arriba, obtenemos la conclusión. □

Método Quasi-Newton y regla de Wolfe

Un detalle importante en el teorema anterior es que el paso α_k para el Método de Newton-Raphson (6.8) se toma constante e igual a 1. Veremos ahora que si la dirección de descenso del Método Quasi-Newton es una buena aproximación de la del Método de Newton-Raphson, entonces el paso $\alpha_k=1$ es admisible para la regla de Wolfe y el método converge de forma cuadrática.

Teorema 6.6. *Sea $f\colon \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ una función propia y dos veces Gâteaux diferenciable en $dom(f)$, el cual asumimos ser un abierto de $\mathbb{R}^n$. Supongamos que existe $\bar{x}\in\mathbb{R}^n$ tal que $\nabla f(\bar{x})=0$ y $\nabla^2 f(\bar{x})\in\mathbb{S}^n_{++}(\mathbb{R})$, y que además $\nabla^2 f$ es localmente Lipschitz continua en torno a $\bar{x}$. Sea $x_0\in\mathbb{R}^n$ y consideremos la sucesión generada por la recurrencia*

$$x_{k+1}=x_k-\alpha_k B_k^{-1}\nabla f(x_k),\qquad \forall k\in\mathbb{N},$$

con $\{\alpha_k\}$ dado por la regla de Wolfe (6.6) con $\omega_1\in(0,1/2)$. Entonces existe $\rho>0$ tal que

1. *Si $\alpha_k=1$ para todo $k\in\mathbb{N}$, $x_0\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$ y $\|B_k-\nabla^2 f(\bar{x})\|\le\rho$, entonces $\{x_k\}$ converge a $\bar{x}$ linealmente.*

2. *Si además se satisface*

$$\lim_{k\to+\infty}\frac{|(B_k-\nabla^2 f(\bar{x}))d_k|}{|d_k|}=0, \tag{6.9}$$

   entonces

$$\lim_{k\to\infty}\frac{|x_{k+1}-\bar{x}|}{|x_k-\bar{x}|}=\lim_{k\to\infty}\frac{|\nabla f(x_{k+1})|}{|\nabla f(x_k)|}=0.$$

3. *Existe $k_0\in\mathbb{N}$ tal que el paso $\alpha_k=1$ satisface la regla de Wolfe para todo $k\ge k_0$.*

Demostración. Recordemos que, para $x\in\mathrm{dom}(f)$ habíamos denotado por λ_x al menor valor propio de $\nabla^2 f(x)$. Como $\nabla^2 f(\bar{x})\in\mathbb{S}^n_{++}(\mathbb{R})$, de la Proposición 4.2 se tiene

$$y^\top\nabla^2 f(\bar{x})y\ge\lambda_{\bar{x}}|y|^2,\quad\forall y\in\mathbb{R}^n,$$

donde $\lambda_{\bar{x}}>0$. Para todo $y\in\mathbb{R}^n$ y $k\in\mathbb{N}$ se tiene

$$\begin{aligned} y^\top B_k y &= y^\top\nabla^2 f(\bar{x})y+y^\top(B_k-\nabla^2 f(\bar{x}))y\\ &\ge\lambda_{\bar{x}}|y|^2-\|B_k-\nabla^2 f(\bar{x})\||y|^2. \end{aligned} \tag{6.10}$$

Luego, definiendo $\rho=\lambda_{\bar{x}}\min\{1/8,1/(4L)\}$, si $\|B_k-\nabla^2 f(\bar{x})\|\le\rho$ se tiene que B_k es definida positiva, para todo $y\in\mathbb{R}^n$, $y^\top B_k y\ge 7\lambda_{\bar{x}}|y|^2/8$ y existen matrices P_k y D_k tales que $B_k=P_kD_kP_k^\top$ con $P_k^{-1}=P_k^\top$, de modo que $B_k^{-1}=P_kD_k^{-1}P_k^\top$ y

$$\|B_k^{-1}\|\le\frac{8}{7\lambda_{\bar{x}}}\le\frac{2}{\lambda_{\bar{x}}}.$$

Supongamos que $x_k\in B_{\mathbb{R}^n}(\bar{x},\rho)$ para algún $k\in\mathbb{N}$. Para simplificar la notación, notemos $g_k=\nabla f(x_k)$. De (6.8) se deduce que si $x_k=\bar{x}$ entonces $x_{k+1}=\bar{x}$, por lo que suponemos que $x_k\ne\bar{x}$. Como $\bar{x}$ es un punto crítico de f, es decir, $\nabla f(\bar{x})=0$, usando la propiedad de Lipschitz continuidad de $\nabla^2 f$ y la relación

$$g_k=\nabla f(x_k)-\nabla f(\bar{x})=\int_0^1\nabla^2 f(\bar{x}+t(x_k-\bar{x}))(x_k-\bar{x})dt,$$

tenemos que

$$\begin{aligned}
|x_{k+1}-\bar{x}| &= |x_k-\bar{x}-B_k^{-1}g_k| \\
&= |B_k^{-1}\left(B_k(x_k-\bar{x})-g_k\right)| \\
&= \left|B_k^{-1}\left(\int_0^1 [B_k-\nabla^2 f(\bar{x}+t(x_k-\bar{x}))](x_k-\bar{x})dt\right)\right| \\
&\leq \frac{2}{\lambda_{\bar{x}}}\left(|(B_k-\nabla^2 f(\bar{x}))(x_k-\bar{x})|+|x_k-\bar{x}|\int_0^1 \|\nabla^2 f(\bar{x})-\nabla^2 f(\bar{x}+t(x_k-\bar{x}))\|dt\right) \\
&\leq \frac{2}{\lambda_{\bar{x}}}|x_k-\bar{x}|\left(\rho+L|x_k-\bar{x}|\int_0^1 tdt\right) \\
&\leq \frac{1}{2}|x_k-\bar{x}|,
\end{aligned}$$

y luego $x_{k+1}\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$. Además, usando inducción vemos que la sucesión $\{x_k\}$ está contenida en $\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$ si $x_0\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$ y

$$|x_{k+1}-\bar{x}|\leq\frac{1}{2^{k+1}}|x_0-\bar{x}|,\quad \forall k\in\mathbb{N}.$$

De aquí se concluye que $x_k\to\bar{x}$ y la convergencia es lineal. Por otra parte, como $x_k\in\mathbb{B}_{\mathbb{R}^n}(\bar{x},\rho)$, $\|\nabla^2 f(x_k)-\nabla^2 f(\bar{x})\|\leq L|x_k-\bar{x}|\leq\lambda_{\bar{x}}/4$ y luego argumentando como en (6.10) se deduce $\|\nabla^2 f(x_k)^{-1}\|\leq 4/(3\lambda_{\bar{x}})\leq 2/\lambda_{\bar{x}}$ y

$$\begin{aligned}
|x_{k+1}-\bar{x}| &= |x_k-\bar{x}-B_k^{-1}g_k| \\
&\leq |x_k-\bar{x}-\nabla^2 f(x_k)^{-1}g_k|+|(B_k^{-1}-\nabla^2 f(x_k)^{-1})g_k| \\
&= |\nabla^2 f(x_k)^{-1}\left(\nabla^2 f(x_k)(x_k-\bar{x})-g_k\right)|+|\nabla^2 f(x_k)^{-1}(\nabla^2 f(x_k)-B_k)d_k| \\
&\leq |\nabla^2 f(x_k)^{-1}|\left(\left|\int_0^1 [\nabla^2 f(x_k)-\nabla^2 f(\bar{x}+t(x_k-\bar{x}))](x_k-\bar{x})dt\right|+|(\nabla^2 f(x_k)-B_k)d_k|\right) \\
&\leq \frac{2}{\lambda_{\bar{x}}}\left(|x_k-\bar{x}|\int_0^1 \|\nabla^2 f(x_k)-\nabla^2 f(\bar{x}+t(x_k-\bar{x}))\|dt+|(\nabla^2 f(x_k)-B_k)d_k|\right) \\
&\leq \frac{2}{\lambda_{\bar{x}}}\left(\frac{L}{2}|x_k-\bar{x}|^2+|(\nabla^2 f(\bar{x})-B_k)d_k|+\|(\nabla^2 f(x_k)-\nabla^2 f(\bar{x}))\|\,|d_k|\right) \\
&\leq \frac{2}{\lambda_{\bar{x}}}\left(\frac{L}{2}|x_k-\bar{x}|^2+|(\nabla^2 f(\bar{x})-B_k)d_k|+L|x_k-\bar{x}|\,|d_k|\right). \qquad (6.11)
\end{aligned}$$

Notando que (6.9) asegura la existencia de $k_0\in\mathbb{N}$ tal que, para todo $k\geq k_0$,

$$\frac{|(\nabla^2 f(\bar{x})-B_k)d_k|}{|d_k|}\leq\rho\leq\frac{\lambda_{\bar{x}}}{8},$$

se tiene que, para todo $k\geq k_0$,

$$\begin{aligned}
\frac{|d_k|}{|x_k-\bar{x}|} &\leq \frac{|x_{k+1}-\bar{x}|+|x_k-\bar{x}|}{|x_k-\bar{x}|} \\
&= 1+\frac{L}{\lambda_{\bar{x}}}|x_k-\bar{x}|+\frac{2}{\lambda_{\bar{x}}}\frac{|(\nabla^2 f(\bar{x})-B_k)d_k|}{|d_k|}\frac{|d_k|}{|x_k-\bar{x}|}+\frac{2L}{\lambda_{\bar{x}}}|x_k-\bar{x}|\frac{|d_k|}{|x_k-\bar{x}|} \\
&\leq 1+\frac{L}{\lambda_{\bar{x}}}|x_k-\bar{x}|+\frac{3}{4}\frac{|d_k|}{|x_k-\bar{x}|},
\end{aligned}$$

y, por lo tanto, para todo $k \geq k_0$,

$$\frac{|d_k|}{|x_k - \bar{x}|} \leq 4 + \frac{4L}{\lambda_{\bar{x}}}|x_k - \bar{x}|.$$

Luego, de (6.11) se deduce

$$\frac{|x_{k+1} - \bar{x}|}{|x_k - \bar{x}|} \leq \frac{2}{\lambda_{\bar{x}}}\left(\frac{L}{2}|x_k - \bar{x}| + \frac{|(\nabla^2 f(\bar{x}) - B_k)d_k|}{|d_k|}\frac{|d_k|}{|x_k - \bar{x}|} + L|x_k - \bar{x}|\frac{|d_k|}{|x_k - \bar{x}|}\right) \to 0$$

cuando $k \to \infty$ y se deduce la convergencia superlineal. Por otra parte, dado que $B_k(x_{k+1} - x_k) + g_k = 0$ para todo $k \in \mathbb{N}$, tenemos que

$$\begin{aligned}
|g_{k+1}| &= |g_{k+1} - g_k - B_k(x_{k+1} - x_k)| \\
&= \left|\int_0^1 \nabla^2 f(x_k + t(x_{k+1} - x_k))(x_{k+1} - x_k)dt - B_k(x_{k+1} - x_k)\right| \\
&\leq |d_k|\int_0^1 \|(\nabla^2 f(x_k + t(x_{k+1} - x_k)) - \nabla^2 f(\bar{x}))\|dt + |(B_k - \nabla^2 f(\bar{x}))d_k| \\
&\leq L|d_k|\left(|x_k - \bar{x}| + \frac{1}{2}|d_k|\right) + |(B_k - \nabla^2 f(\bar{x}))d_k| \\
&= |d_k|\left(L\left(|x_k - \bar{x}| + \frac{1}{2}|d_k|\right) + \frac{|(B_k - \nabla^2 f(\bar{x}))d_k|}{|d_k|}\right) \\
&\leq \|B_k^{-1}\||g_k|\left(L\left(|x_k - \bar{x}| + \frac{1}{2}|d_k|\right) + \frac{|(B_k - \nabla^2 f(\bar{x}))d_k|}{|d_k|}\right),
\end{aligned}$$

de donde $|g_{k+1}|/|g_k| \to 0$ cuando $k \to \infty$.

Ahora probemos que, para todo $k \geq k_0$, $\alpha_k = 1$ satisface la regla de Wolfe (6.6). De hecho, dado $k \geq k_0$ y usando la expansión de orden 2 para $f(x_k + d_k)$ en torno a $d_k = 0$, se tiene

$$f(x_k + d_k) = f(x_k) + \nabla f(x_k)^\top d_k + \frac{1}{2}d_k^\top \nabla^2 f(x_k)d_k + o(|d_k|^2), \tag{6.12}$$

y de $d_k = -B_k^{-1}\nabla f(x_k)$ se obtiene

$$\begin{aligned}
f(x_k + d_k) - f(x_k) - \omega_1 \nabla f(x_k)^\top d_k &= (1 - \omega_1)\nabla f(x_k)^\top d_k + \frac{1}{2}d_k^\top \nabla^2 f(x_k)d_k + o(|d_k|^2) \\
&= -(1 - \omega_1)d_k^\top B_k d_k + \frac{1}{2}d_k^\top \nabla^2 f(x_k)d_k + o(|d_k|^2) \\
&= (1 - \omega_1)d_k^\top(\nabla^2 f(\bar{x}) - B_k)d_k + \frac{1}{2}d_k^\top(\nabla^2 f(x_k) - \nabla^2 f(\bar{x}))d_k \\
&\quad - (1/2 - \omega_1)d_k^\top \nabla^2 f(\bar{x})d_k + o(|d_k|^2) \\
&\leq (1 - \omega_1)|d_k||(\nabla^2 f(\bar{x}) - B_k)d_k| + \frac{1}{2}|d_k|^2\|\nabla^2 f(x_k) - \nabla^2 f(\bar{x})\| \\
&\quad - (1/2 - \omega_1)\lambda_{\bar{x}}|d_k|^2 + o(|d_k|^2),
\end{aligned}$$

donde $\lambda_x > 0$ es el menor valor propio de $\nabla^2 f(x)$. Dividiendo por $|d_k|^2$, usando (6.9) y la continuidad de $\nabla^2 f$ se tiene que la primera condicion de Wolfe (6.6) se satisface con $\omega_1 \in (0, 1/2)$. Para la segunda condición, por teorema del valor medio se tiene que existe $\lambda \in (0,1)$ tal que

$$\nabla f(x_k + d_k)^\top d_k - \nabla f(x_k)^\top d_k = d_k^\top \nabla^2 f(x_k + \lambda d_k)d_k$$

y luego

$$\begin{aligned}\nabla f(x_k+d_k)^\top d_k - \omega_2 \nabla f(x_k)^\top d_k &= (1-\omega_2)\nabla f(x_k)^\top d_k + d_k^\top \nabla^2 f(x_k+\lambda d_k) d_k \\ &= -(1-\omega_2) d_k^\top B_k d_k + d_k^\top \nabla^2 f(x_k+\lambda d_k) d_k \\ &= (1-\omega_2) d_k^\top (\nabla^2 f(\bar{x}) - B_k) d_k + d_k^\top \left(\nabla^2 f(x_k+\lambda d_k) - \nabla^2 f(\bar{x})\right) d_k \\ &\quad + \omega_2 d_k^\top \nabla^2 f(\bar{x}) d_k \\ &\geq (1-\omega_2) d_k^\top (\nabla^2 f(\bar{x}) - B_k) d_k + d_k^\top \left(\nabla^2 f(x_k+\lambda d_k) - \nabla^2 f(\bar{x})\right) d_k \\ &\quad + \omega_2 \lambda_{\bar{x}} |d_k|^2,\end{aligned}$$

y el resultado se obtiene como antes. □

6.4.5. Fórmula explícitas para Quasi-Newton

Ahora mostraremos algunas formas constructivas de determinar las matrices B_k para el método Quasi-Newton. La primera que veremos se llama fórmula DFP en honor a sus descubridores (Davidon-Fletcher-Powell) y la segunda se llamada fórmula BFGS por sus descubridores (Broyden-Fletcher-Goldfarb-Shanno). Mostraremos en particular que la fórmula BFGS verifica la condición (6.13), lo que asegura la convergencia cuadrática del método al tomar paso $\alpha_k = 1$ para todo $k \in \mathbb{N}$ suficientemente grande (gracias al Teorema 6.6).

Preliminares

Describamos la idea esencial de ambos métodos. Supongamos conocida la iteración del Método Quasi-Newton $x_k \in \mathbb{R}^n$ y la matriz $B_k \in \mathbb{S}^n_{++}(\mathbb{R})$. Consideremos la función $m_k : \mathbb{R}^n \to \mathbb{R}$ dada por

$$m_k(d) = f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2} d^\top B_k d, \qquad \forall d \in \mathbb{R}^n.$$

Esta función tiene la propiedad que $m_k(0) = f(x_k)$ y $\nabla m_k(0) = \nabla f(x_k)$. Además, al ser B_k simétrica y definida positiva tenemos que m_k es coerciva y por lo tanto tiene un único mínimo, digamos d_k, que está caracterizado por la regla de Fermat. Dado que B_k es invertible, no es difícil ver que d_k está dado por la fórmula

$$d_k = -B_k^{-1} \nabla f(x_k), \quad \forall k \in \mathbb{N}. \tag{6.13}$$

Es decir, es la dirección dada por el Método Quasi-Newton. Ahora bien, si tuviésemos a disposición la siguiente iteración del Método Quasi-Newton x_{k+1}, nos gustaría hacer algo similar para determinar d_{k+1}. Para esto, definimos la función $f_{k+1} : \mathbb{R}^n \to \mathbb{R}$ dada por

$$f_{k+1}(x) = f(x_{k+1}) + \nabla f(x_{k+1})^\top (x - x_{k+1}) + \frac{1}{2}(x - x_{k+1})^\top B_{k+1}(x - x_{k+1}), \qquad \forall x \in \mathbb{R}^n.$$

Es claro que $\nabla f_{k+1}(x_{k+1}) = \nabla f(x_{k+1})$. Nos gustaría además que f_{k+1} fuese también una buena aproximación de f, para esto podemos pedir por ejemplo que $\nabla f_{k+1}(x_k) = \nabla f(x_k)$, lo que se traduce en:

$$B_{k+1} s_k = y_k, \qquad \text{con } s_k = x_{k+1} - x_k = \alpha_k d_k \text{ e } y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

Esta última, se conoce como la *ecuación de la secante*; notar que la incógnita en este caso es la matriz B_{k+1}. Ahora bien, dado que buscamos que B_{k+1} sea definida positiva, necesitamos que $s_k^\top B_{k+1} s_k > 0$. Luego para que la ecuación de la secante tenga solución necesitamos que $s_k^\top y_k > 0$. Esto se puede asegurar si por ejemplo α_k satisface la condición de Wolfe (6.6b). Efectivamente, si $\alpha_k > 0$ se escoge usando la regla de Wolfe tendremos que

$$s_k^\top y_k = \alpha_k d_k^\top y_k \geq \alpha_k(\omega_2 - 1) d_k^\top \nabla f(x_k) > 0.$$

Ahora bien, dado que la ecuación de la secante es una ecuación matricial, ésta posee infinitas soluciones pues esta ecuación se compone de n ecuaciones que sumado a las n desigualdades provenientes del hecho que B_{k+1} es definida positiva, no compensan los $\frac{1}{2}n(n+1)$ grados de libertad de la simetría de B_{k+1}.

Formula DFP

Una forma de construir B_{k+1} es buscando, entre todas las soluciones a la ecuación de la secante, la matriz más próxima a B_k en algún sentido. Dicho de otra forma, B_{k+1} será la proyección B_k sobre el espacio de soluciones de la ecuación de la secante. Esto se puede formular como el siguiente problema de optimización

$$(\mathrm{P_{DFP}}) \qquad \text{Minimizar } \|B - B_k\| \quad \text{sobre todos los } B \in \mathbb{S}^n(\mathbb{R}) \text{ tales que } Bs_k = y_k,$$

donde $s_k^\top y_k > 0$, $B_k \in \mathbb{S}^n_{++}(\mathbb{R})$ y $M \mapsto \|M\|$ es una norma sobre $\mathbb{S}^n(\mathbb{R})$.

Observación 6.2. *Para cada norma utilizada se obtendrá un forma de calcular B_{k+1} y por lo tanto un nuevo Método Quasi-Newton.*

La fórmula DFP utiliza la norma

$$\|M\| = \sqrt{\mathrm{tr}(W^{1/2} M W M W^{1/2})} \quad \text{con} \quad W = \int_0^1 \nabla^2 f(x_k + t\alpha_k d_k)\, dt.$$

La matriz W se conoce como la *matriz Hessiana promedio* de f y no es difícil ver que, gracias al teorema fundamental del cálculo, W es una solución particular de la ecuación de la secante.

Bajo estas condiciones y usando las condiciones de optimalidad del problema ($\mathrm{P_{DFP}}$), se tiene que la matriz B_{k+1} queda determinada por la recurrencia:

$$B_{k+1} = \left(I - \frac{1}{y_k^\top s_k} y_k s_k^\top\right) B_k \left(I - \frac{1}{y_k^\top s_k} s_k y_k^\top\right) + \frac{1}{y_k^\top s_k} y_k y_k^\top, \quad \forall k \in \mathbb{N}.$$

Ahora bien, en el Método Quasi-Newton nos interesa conocer la inversa de B_k y no necesariamente B_k misma. Dada la estructura de B_{k+1}, podemos calcular su inversa usando la fórmula de Sherman-Morrison-Woodbury:

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u} \quad \forall A \in \mathbb{M}_{n\times n}(\mathbb{R}) \text{ invertible}, \forall u, v \in \mathbb{R}^n.$$

Esto implica que la fórmula DFP está dada por:

$$(\mathrm{DFP}) \qquad B_{k+1}^{-1} = B_k^{-1} - \frac{1}{y_k^\top B_k^{-1} y_k} B_k^{-1} y_k y_k^\top B_k^{-1} + \frac{1}{y_k^\top s_k} s_k s_k^\top, \quad \forall k \in \mathbb{N}.$$

Formula BFGS

Una forma alternativa de obtener un método Quasi-Newton es calculando directamente la inversa y plantear el problema (P_{DFP}) de una forma equivalente pero para la inversa de B_{k+1}. En términos de problema de optimización esto se escribe como sigue

$$(P_{BFGS}) \qquad \text{Minimizar } \|M - B_k^{-1}\| \quad \text{sobre todos los } M \in \mathbb{S}^n(\mathbb{R}) \text{ tales que } My_k = s_k,$$

donde $s_k^\top y_k > 0$, $B_k \in \mathbb{S}^n_{++}(\mathbb{R})$ y $M \mapsto \|M\|$ es una norma sobre $\mathbb{S}^n(\mathbb{R})$. Notar que en este caso se tiene que M^{-1} será solución de la ecuación de la secante. Luego, usando las condiciones de optimalidad del problema (P_{BFGS}), se tiene que la matriz B_{k+1} queda determinada por la recurrencia:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$

y por lo tanto la fórmula BFGS viene dada por

$$(BFGS) \qquad B_{k+1}^{-1} = \left(I - \frac{1}{y_k^\top s_k} s_k y_k^\top\right) B_k^{-1} \left(I - \frac{1}{y_k^\top s_k} y_k s_k^\top\right) + \frac{1}{y_k^\top s_k} s_k s_k^\top, \quad \forall k \in \mathbb{N}.$$

Veamos ahora un teorema sobre la convergencia global del Método Quasi-Newton usando la fórmula BFGS. Cabe destacar que bajo las hipótesis del siguiente resultado, la función objetivo es coerciva y por lo tanto tiene un mínimo.

Teorema 6.7. *Sea $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función propia y dos veces Gâteaux diferenciable en $dom(f)$, el cual asumimos ser un abierto de $\mathbb{R}^n$. Supongamos que $x_0 \in \mathbb{R}^n$ es tal que $\Gamma_{f(x_0)}(f)$ es convexo y existen $\lambda, \sigma > 0$ tal que*

$$\lambda |y|^2 \leq y^\top \nabla^2 f(x) y \leq \sigma |y|^2, \quad \forall x \in \Gamma_{f(x_0)}(f),\ \forall y \in \mathbb{R}^n.$$

Entonces, la secuencia $\{x_k\}$ generada por el Método Quasi-Newton, con B_k determinada por la fórmula BFGS, con paso α_k dado por la regla de Wolfe (6.6) converge a $\bar{x} \in \arg\min_{\mathbb{R}^n}(f)$.

Finalmente veremos que la tasa de convergencia del Método Quasi-Newton es cuadrática si las matrices B_k se escogen usando la fórmula BFGS.

Teorema 6.8. *Sea $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función propia y dos veces Gâteaux diferenciable en $dom(f)$, el cual asumimos ser un abierto de $\mathbb{R}^n$. Supongamos que existe $\bar{x} \in \mathbb{R}^n$ tal que $\nabla f(\bar{x}) = 0$ y $\nabla^2 f(\bar{x}) \in \mathbb{S}^n_{++}(\mathbb{R})$, y que además $\nabla^2 f$ es localmente Lipschitz continua en torno a $\bar{x}$. Supongamos que el método BFGS converge al punto crítico $\bar{x}$. Luego, si*

$$\sum_{k=0}^{\infty} |x_k - \bar{x}| < +\infty,$$

entonces x_k converge a $\bar{x}$ a una tasa superlineal, es decir,

$$\lim_{k\to\infty} \frac{|(B_k - \nabla^2 f(\bar{x}))(x_{k+1} - x_k)|}{|x_{k+1} - x_k|} = 0$$

6.5. Ejercicios

1. MÍNIMOS LOCALES QUE SON GLOBALES

   Sea $(\mathbf{X}, \|\cdot\|)$ un espacio vectorial normado y $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ continua en $\text{dom}(f)$. Muestre que $\bar{x}$ es un mínimo global de f si y sólo si todo x tal que $f(x) = f(\bar{x})$ es un mínimo local de f.

2. MAXIMIZACIÓN DE UTILIDADES

   Una pesquera maneja dos variables en su proceso de extracción mensual, la cantidad de horas-hombre utilizada (variable x) y la superficie que se abarca (variable y), la cuales (debido a las unidades en que se miden) satisfacen que $x > 0$ e $y > 1$. Así, dados dos valores x e y para estas variables, la cosecha mensual (en kilos) está dada por:

   $$\text{cosecha} = x^{\alpha} \log^{\beta}(y),$$

   donde α y β son dos parámetros dados. Si el precio del kilo de pescado es $p = 1$, y los costos unitarios asociados a x e y son los valores estrictamente positivos c_x y c_y, respectivamente, entonces:

   a) Modele el problema de maximizar el beneficio de la pesquera como uno de programación sin restricciones en $x > 0$ e $y > 1$, y encuentre las relaciones de la forma $h(y) = 0$ e $x = g(y)$ que satisfacen los puntos críticos del problema. ¿Puede concluir que estos son efectivamente máximos?

   b) Desde ahora sabemos que los parámetros satisfacen $\alpha \in [0, 1)$ y $\beta \geq 0$, y reducimos nuestra estrategia al conjunto

   $$\mathbf{S} := \left\{ (x, y) \in \mathbb{R}^2 \;\middle|\; x > 0,\ y > 1,\ \log(y) > \frac{\beta}{1-\alpha} - 1 \right\}.$$

   Demuestre que si los puntos críticos de la parte anterior están en $\mathbf{S}$, entonces estos son máximos (globales) del problema.

   Indicación: Estudie la convexidad del negativo de la función de beneficios.

3. Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función p veces continuamente diferenciable en el interior de su dominio (con $p \geq 2$), tal que para $\bar{x} \in \text{int}(\text{dom}\, f)$ se tiene:

   $$\nabla^i f(\bar{x}) = 0,\ \forall i = 1, ..., p-1 \quad \text{y} \quad \nabla^p f(\bar{x}) \neq 0.$$

   Demostrar que para que $\bar{x}$ sea un mínimo (local) de f,

   (a) es necesario que p sea par y $\nabla^p f(\bar{x})(h, ..., h) \geq 0$ para todo $h \in \mathbb{R}^n$.

   (b) es suficiente que p sea par y $\nabla^p f(\bar{x})(h, ..., h) > 0$ para todo $h \in \mathbb{R}^n$.

CAPÍTULO 7

Optimización restricta

Abstract. En este capítulo estudiaremos problemas de optimización donde se busca minimizar una función diferenciable sobre un conjunto de restricciones dado. Al igual que en el capítulo anterior, el problema que enfrentaremos no será necesariamente convexo. Estudiaremos las condiciones de optimalidad (necesarias y suficientes) para que un punto sea un mínimo local y estudiaremos algunos métodos iterativos para encontrar mínimos locales. Pondremos particular énfasis en el problema de Programación Matemática.

En esta parte, al igual que en el capítulo anterior, usaremos la intuición desarrollada para la optimización convexa con restricciones para estudiar problemas generales de optimización con restricciones. En particular nos enfocaremos en restricciones que se pueden escribir como intersecciones de de variedades y conjuntos de subnivel inferiores. Esta clase de problemas recibe el nombre de problemas de *Programación Matemática*.

A lo largo de este capítulo, trabajaremos básicamente con funciones que son localmente Lipschitz continuas y Gâteaux diferenciable en el interior de sus dominios. La primera parte de la exposición se hará para un espacio vectorial normado arbitrario $\mathbf{X}$, pero la parte de Programación Matemática será sobre espacios de Hilbert (de dimensión finita en algunos casos, pero no necesariamente $\mathbb{R}^n$).

7.1. Problema de Optimización No Lineal General

En esta parte nos enfocaremos en el problema general de optimización

(P) Minimizar $f(x)$ sobre todos los $x \in \mathbf{X}$ que satisfacen la restricción $x \in \mathbf{S}$

donde $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ es una función no lineal general y $\mathbf{S} \subseteq \mathbf{X}$ es conjunto cerrado no vacío. Dado que queremos tratar en adelante el caso general, no necesariamente convexo, la teoría que desarrollaremos será, al igual que en el capítulo anterior, sólo local. Para ello debemos extender la noción de mínimo local para problemas con restricciones.

Definición 7.1 (Mínimos locales)**.** *Sea* $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ *un espacio de Hilbert,* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función dada y* $\mathbf{S} \subseteq \mathbf{X}$ *un conjunto no vacío. Un punto* $\bar{x} \in dom(f) \cap \mathbf{S}$ *se dice mínimo local del problema* (P) *si existe* $r > 0$ *tal que*

$$f(\bar{x}) \leq f(x), \qquad \forall x \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r) \cap \mathbf{S}.$$

Un mínimo local de (P) *se dice* estricto *si la relación anterior es válida con desigualdad estricta.*

Al igual que en el capítulo anterior, para no generar confusión, a los mínimos del problema (P) les agregaremos el adjetivo *global* para distinguirlos de los mínimos locales. De forma similar al caso sin restricciones, todo mínimo global del problema (P) es también un mínimo local y la existencia de mínimos locales no asegura siquiera que la función sea acotada inferiormente. Además, todo punto que pertenece a $\mathbf{S}$ se dirá *factible* para el problema (P).

7.1.1. Condiciones de Optimalidad de primer orden

Recordemos que en el caso convexo, logramos escribir las condiciones de optimalidad usando la noción de cono normal. En otras palabras, mostramos que $\bar{x} \in \operatorname{sol}(P)$ si y sólo si

$$\bar{x} \in \mathbf{S} \quad \text{y} \quad -\nabla f(\bar{x}) \in N_{\mathbf{S}}(\bar{x}) := \{\eta \in \mathbf{X} \mid \langle \eta, x - \bar{x} \rangle \leq 0, \quad \forall x \in \mathbf{S}\}.$$

Ahora veremos una contraparte tangencial esta condición.

Definición 7.2 (Cono Tangente)**.** *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y $\mathbf{S} \subseteq \mathbf{X}$ un conjunto dado. Definimos el* cono tangente *a $\mathbf{S}$ en $x \in \mathbf{S}$ via la fórmula*

$$T_{\mathbf{S}}(x) := \{d \in \mathbf{X} \mid \exists \{(t_k, d_k)\} \subseteq (0, +\infty) \times \mathbf{X} \text{ tal que } (t_k, d_k) \to (0, d) \text{ con } x + t_k d_k \in \mathbf{S}, \ \forall k \in \mathbb{N}\}.$$

Observación 7.1. *No es difícil ver $T_{\mathbf{S}}(x)$ es un cono cerrado para todo $x \in \mathbf{S}$ y que además $T_{\mathbf{S}}(x) = \mathbf{X}$ si $x \in \operatorname{int}(\mathbf{S})$. Más aún, tenemos que*

$$\eta \in N_{\mathbf{S}}(x) \quad \Longrightarrow \quad \langle \eta, d \rangle \leq 0, \quad \forall d \in T_{\mathbf{S}}(x),$$

pero la implicancia recíproca no es necesariamente cierta. En efecto, sea $\mathbf{S} = \{x \in \mathbb{R}^2 \mid x_2 = 0 \vee x = \bar{x}\}$, donde $\bar{x} = (0,1)$. En este caso se tiene que $T_{\mathbf{S}}(0,0) = \{x \in \mathbb{R}^2 \mid x_2 = 0\}$ y por lo tanto para $\eta = \bar{x}$ se tiene

$$\eta^\top d = 0, \quad \forall d \in T_{\mathbf{S}}(0,0),$$

pero $\eta \notin N_{\mathbf{S}}(0,0)$, pues $\eta^\top(\bar{x} - (0,0)) = |\bar{x}|^2 = 1 > 0$. Cabe destacar que la recíproca es cierta si $\mathbf{S}$ es convexo (ver Ejercicio 2). En particular, Teorema 7.1 más abajo es equivalente a Teorema 5.3 bajo hipótesis de convexidad y diferenciabilidad apropiadas.

Con esta herramienta podemos ahora estudiar condiciones de optimalidad para el problema general de Optimización No Lineal.

Teorema 7.1 (Condición Necesaria de Primer Orden)**.** *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert. Sea $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función propia, localmente Lipschitz continua y Gâteaux diferenciable en una vecindad de $\bar{x} \in \mathbf{X}$. Si $\bar{x}$ es un mínimo local de (P), entonces*

$$\text{(CNPO)} \qquad \langle \nabla f(\bar{x}), d \rangle \geq 0, \qquad \forall d \in T_{\mathbf{S}}(\bar{x}).$$

Demostración. Como $\bar{x}$ es mínimo local, existe $r > 0$ tal que $f(\bar{x}) \leq f(x)$ para todo $x \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r) \cap \mathbf{S}$. Dado que f es localmente Lipschitz en una vecindad de $\bar{x}$, sin perdida de generalidad podemos asumir que existe $L > 0$ tal que

$$|f(x) - f(y)| \leq L|x - y|, \qquad \forall x, y \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r).$$

Sea $d \in T_{\mathbf{S}}(\bar{x}) \setminus \{0\}$ (si $d = 0$ la conclusión es directa). Luego, existen sucesiones $\{t_k\} \subseteq (0, +\infty)$ y $\{d_k\} \subseteq \mathbb{R}^n$ tales que $t_k \to 0$, $d_k \to d$ y $\bar{x} + t_k d_k \in \mathbf{S}$ para todo $k \in \mathbb{N}$. Entonces, existe $k_0 \in \mathbb{N}$ tal que

$$\bar{x} + t_k d_k \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r) \cap \mathbf{S} \quad \text{y} \quad \bar{x} + t_k d \in \mathbb{B}_{\mathbf{X}}(\bar{x}, r), \qquad \forall k \geq k_0.$$

En consecuencia, para todo $k \in \mathbb{N}$ con $k \geq k_0$ tenemos que

$$0 \leq \frac{f(\bar{x} + t_k d_k) - f(\bar{x})}{t_k} = \frac{f(\bar{x} + t_k d_k) - f(\bar{x} + t_k d)}{t_k} + \frac{f(\bar{x} + t_k d) - f(\bar{x})}{t_k} \leq L|d_k - d| + \frac{f(\bar{x} + t_k d) - f(\bar{x})}{t_k}.$$

Finalemente, el resultado se obtiene tomando límite $k \to \infty$ y usando que $t_k \to 0$ y $d_k \to d$. □

Notemos que el Teorema 7.1 es una generalización del Teorema 6.1, pues en el caso que no hay restricciones, es decir $\mathbf{S} = \mathbf{X}$, se tiene que $T_{\mathbf{S}}(x) = \mathbf{X}$ para todo $x \in \mathbf{X}$; esto se debe a que $\operatorname{int}(\mathbf{X}) = \mathbf{X}$.

Ejemplo 7.1.1. *Cabe también destacar que el Teorema 7.1 al igual que el Teorema 6.1, es sólo una condición necesaria y puede no ser suficiente. En efecto, consideremos la función* $f(x) = x_1$ *y la restricción* $\mathbf{S} = \{x \in \mathbb{R}^2 \mid \sqrt{|x_1|} \leq x_2\}$*; ver Figura 7.1. Luego, tenemos* $\nabla f(0,0) = (1,0)$ *y además* $T_{\mathbf{S}}(0,0) = \{d \in \mathbb{R}^2 \mid d_1 = 0,\ d_2 \geq 0\}$*. Con esto vemos que* (CNPO) *se satisface en el punto* $\bar{x} = (0,0)$*. Sin embargo, este punto no es mínimo local pues, dado* $\alpha > 0$*, cualquier punto de la forma* $x_\alpha = (-\alpha^2, \alpha) \in \mathbf{S}$ *pertenece a* $\mathbf{S}$ *y satisface* $f(x_\alpha) = -\alpha^2$*. Por lo tanto, para cualquier* $\alpha > 0$ *se tiene que* $f(x_\alpha) < 0$ *y* x_α *puede ser tan cercano a* $(0,0)$ *como queramos.*

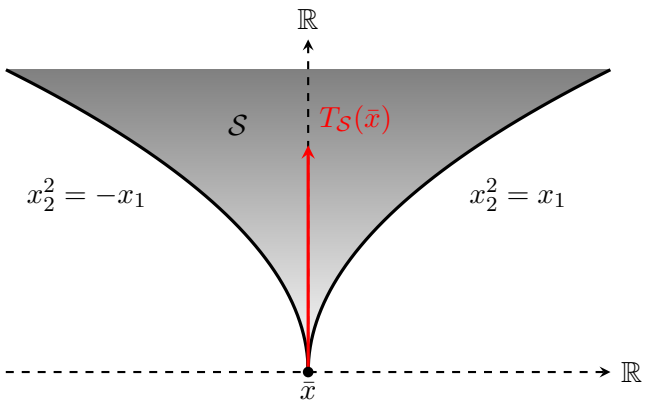


Figura 7.1: Conjunto de restricciones de Ejemplo 7.1.1.

7.2. Programación Matemática

La condición (CNPO) es una condición abstracta que puede ser difícil de manejar, sobre todo porque el cono tangente a un conjunto $\mathbf{S} \subseteq \mathbf{X}$ arbitrario puede ser un objeto complicado a encontrar.

Por esta razón, y para dar un sentido práctico a la condición (CNPO) nos enfocaremos en una clase particular de problemas de optimización, que a su vez es de los más utilizados en aplicaciones. Esta clase de problemas, que llamaremos *Problemas de Programación Matemática*, son aquellos que consisten en minimizar una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ sobre el conjunto de restricciones

$$\mathbf{S} = \left\{x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1,\dots,p, \quad h_j(x) = 0,\ j = 1,\dots,q\right\}.$$

donde $g_1,\dots,g_p : \mathbf{X} \to \mathbb{R}$ y $h_1,\dots,h_q : \mathbf{X} \to \mathbb{R}$ son funciones dadas. En el Capítulo 5, estudiamos un caso particular de este problema (que llamamos Problema de Programación Convexa). A saber, el caso en que las funciones $f, g_1,\dots,g_p$ son convexas y las funciones $h_1,\dots,h_q$ son afines continuas, es decir, para ciertos $x_1^*,\dots,x_q^* \in \mathbf{X}^*$ y $\alpha_1,\dots,\alpha_q \in \mathbb{R}$ se tiene

$$h_j(x) = \langle x_j^*, x\rangle - \alpha_j, \quad \forall j = 1,\dots,q, \qquad \forall x \in \mathbf{X}.$$

A partir de ahora, $\mathbf{X}$ será un espacio de Hilbert dotado de un producto interno denotado $\langle\cdot,\cdot\rangle$. Los ejemplos modelos serán $\mathbf{X} = \mathbb{R}^n$ y $\mathbf{X} = \mathbb{S}^n(\mathbb{R})$.

7.2.1. Cono Linealizante

En el caso convexo vimos que bajo ciertas hipótesis de calificación podíamos dar una expresión explícita para el cono normal al conjunto de restricciones del problema de programación convexa. Ahora nos enfocaremos en obtener algo similar para el caso de la programación matemática.

Definición 7.3. *Sean* $g_1,\dots,g_p : \mathbf{X} \to \mathbb{R}$ *y* $h_1,\dots,h_q : \mathbf{X} \to \mathbb{R}$ *funciones Gâteaux diferenciable y sea*

$$\mathbf{S} = \left\{x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1,\dots,p, \quad h_j(x) = 0,\ j = 1,\dots,q\right\}.$$

Definimos el cono linealizante *a* $\mathbf{S}$ *en* $x \in \mathbf{S}$ *como el conjunto*

$$L_{\mathbf{S}}(x) := \left\{d \in \mathbf{X} \mid \ \langle \nabla g_i(x), d\rangle \leq 0,\ \forall i \in I(x),\ \left\langle \nabla h_j(x), d\right\rangle = 0,\ \forall j \in \{1,\dots,q\}\right\},$$

donde $I(x) = \{i \in \{1,\dots,p\} \mid \ g_i(x) = 0\}$ *es el conjunto de índices de* restricciones activas *en* $x \in \mathbf{S}$.

Notar que el cono linealizante puede ser calculado explícitamente usando los datos del problema, y para ello basta solo conocer las derivadas de las funciones que definen al conjunto de restricciones del problema de programación matemática. Por esta razón, nos gustaría poder escribir (CNPO) usando el cono linealizante en vez del cono tangente. Notemos que en general se tiene que el cono tangente $T_{\mathbf{S}}(x)$ está contenido en el cono linealizante $L_{\mathbf{S}}(x)$, y que esta inclusión puede ser estricta.

Ejemplo 7.2.1. *Sea* $\bar{x} = (1,0)$ *y consideremos el conjunto*

$$\mathbf{S} = \left\{x \in \mathbb{R}^2 \mid x_2 \leq (1-x_1)^3, \quad x_1 \geq 0, \quad y \quad x_2 \geq 0\right\}.$$

Notemos que la primera y tercera restricciones son activas, pero la segunda no. Luego, el cono linealizante al conjunto en $\bar{x}$ *está dado por* $L_{\mathbf{S}}(\bar{x}) = \mathbb{R} \times \{0\}$, *pero* $T_{\mathbf{S}}(\bar{x}) = (-\infty, 0] \times \{0\}$.

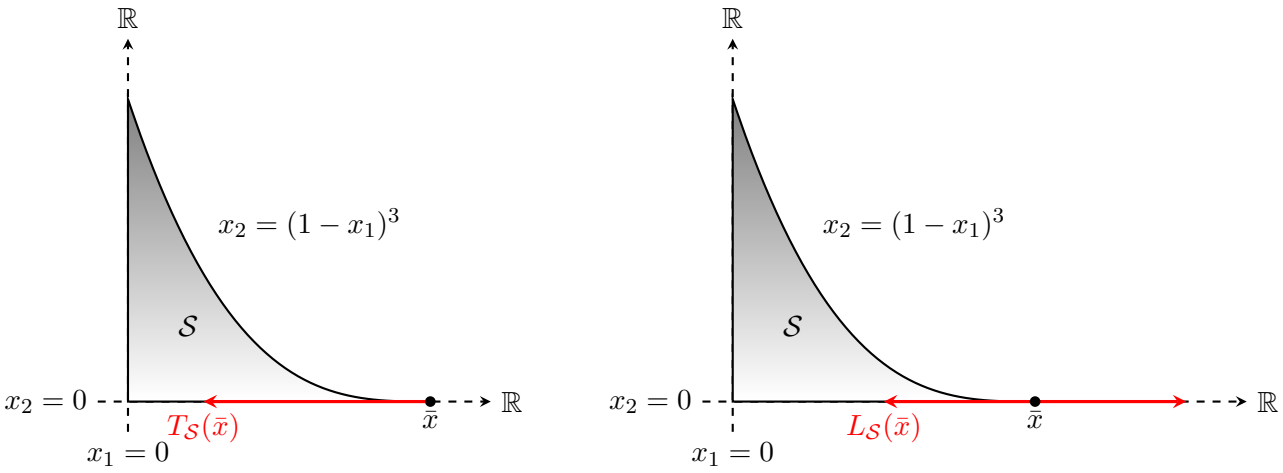


Figura 7.2: Cono tangente y linealizante de Ejemplo 7.2.1.

7.2.2. Condiciones de Calificación

El Ejemplo 7.2.1 muestra que el cono linealizante no coincide necesariamente con el cono tangente, y por lo tanto (CNPO), podría fallar si reemplazásemos indiscriminadamente el cono tangente por el linealizante, pues estaríamos agregando más direcciones de las que necesitamos para estudiar el crecimiento de la función objetivo. Ahora nos enfocaremos en criterios que nos permitirán afirmar que ambos conos, el tangente y linealizante coinciden.

Recuerdo : Funciones continuamente diferenciables

Una función $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ definida en un espacio de Hilbert $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ se dice continuamente diferenciable en $x \in \text{int}(\text{dom}(f))$ si f es Fréchet diferenciable en una vecindad de x y $\nabla f : \mathbf{X} \to \mathbf{X}$ es continuo en una vecindad de x, es decir,

$$\forall \varepsilon > 0,\ \exists r > 0 \text{ tal que } \forall y \in \mathbf{X} \quad \|x - y\| < r \quad \Longrightarrow \quad \|\nabla f(x) - \nabla f(y)\| < \varepsilon.$$

En el caso $\mathbf{X} = \mathbb{R}^n$, esto se reduce a que las derivadas parciales de f sean todas funciones continuas en una vecindad de x. Si $F : \mathbb{R}^n \to \mathbb{R}^m$ es una función vectorial con $F = (F_1, \ldots, F_m)$, esta se dirá continuamente diferenciable si cada función componente $y \mapsto F_i(y)$ es continuamente diferenciable en torno a x.

El Teorema de la Función Implícita es una herramienta fundamental en el Cálculo Diferencial, que dice básicamente que si la ecuación $\Phi(0,u) = 0$ tiene una solución, digamos $\bar{u} \in \mathbb{R}^q$, donde $\Phi : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}^q$ es un campo vectorial dado, entonces se puede construir una curva $u : \mathbb{R} \to \mathbb{R}^q$ que pasa al instante $t = 0$ por $\bar{u}$, tal que

$$\Phi(t, u(t)) = 0, \qquad \forall t \in \mathbb{R} \text{ en una vecindad } t = 0.$$

Recordemos que $J_\Phi(t,u)$ denota la matriz Jacobiana de Φ en el punto (t,u). En este caso, esta matriz tiene la estructura

$$J_\Phi(t,u) = \begin{bmatrix} \partial_t \Phi(t,u) & \nabla_u \Phi(t,u) \end{bmatrix}$$

donde

$$\partial_t \Phi(t,u) := \begin{pmatrix} \partial_t \Phi_1(t,u) \\ \vdots \\ \partial_t \Phi_q(t,u) \end{pmatrix} \quad \text{y} \quad \nabla_u \Phi(t,u) := \begin{pmatrix} \partial_{u_1} \Phi_1(t,u) & \ldots & \partial_{u_q} \Phi_1(t,u) \\ \vdots & \ddots & \vdots \\ \partial_{u_1} \Phi_q(t,u) & \ldots & \partial_{u_q} \Phi_q(t,u) \end{pmatrix}$$

Recuerdo : Teorema de la Función Implícita

Teorema 7.2. *Sea $\Phi : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}^q$ un campo vectorial dado y $\bar{u} \in \mathbb{R}^q$ tal que $\Phi(0,\bar{u}) = 0$. Supongamos que Φ es continuamente diferenciable en una vecindad de $(0,\bar{u})$ con $\nabla_u \Phi(0,\bar{u})$ invertible. Entonces existe $\varepsilon > 0$ y una curva $u : (-\varepsilon, \varepsilon) \to \mathbb{R}^q$, continuamente diferenciable tal que*

$$\Phi(t, u(t)) = 0, \quad \forall t \in (-\varepsilon, \varepsilon) \quad \text{con } u(0) = \bar{u}.$$

Condición de Mangasarian-Fromovitz

La condición de calificación de Mangasarian-Fromovitz (MF) es una de las más utilizadas pues no se considera ser una hipótesis muy exigente para un problema de optimización.

Definición 7.4. *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert. Sean* $g_1,\dots,g_p : \mathbf{X} \to \mathbb{R}$ *y* $h_1,\dots,h_q : \mathbf{X} \to \mathbb{R}$ *funciones Gâteaux diferenciable y sea*

$$\mathbf{S} = \left\{x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1,\dots,p, \quad h_j(x) = 0,\ j = 1,\dots,q\right\}.$$

Diremos que un punto $\bar{x} \in \mathbf{S}$ *satisface la condición de* Mangasarian-Fromovitz *si*

$$\text{(MF)} \qquad \begin{cases} (i) & \left\{\nabla h_1(\bar{x}),\dots,\nabla h_q(\bar{x})\right\} \text{ son linealmente independientes.} \\ (ii) & \exists \bar{d} \in \mathbf{X} \text{ tal que } \left\langle \nabla g_i(\bar{x}), \bar{d}\right\rangle < 0,\ \forall i \in I(\bar{x}) \text{ y } \left\langle \nabla h_j(\bar{x}), \bar{d}\right\rangle = 0,\ \forall j \in \{1,\dots,q\} \end{cases}$$

Esta definición nos permitirá probar que el cono linealizante y el tangente coinciden en todos los puntos que satisfacen (MF). Esto a su vez, es una consecuencia del Teorema de la Función Implícita.

Teorema 7.3. *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert. Sean* $g_1,\dots,g_p : \mathbf{X} \to \mathbb{R}$ *y* $h_1,\dots,h_q : \mathbf{X} \to \mathbb{R}$ *funciones continuamente diferenciables y sea*

$$\mathbf{S} = \left\{x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1,\dots,p, \quad h_j(x) = 0,\ j = 1,\dots,q\right\}.$$

Entonces, $T_{\mathbf{S}}(\bar{x}) \subseteq L_{\mathbf{S}}(\bar{x})$ *para todo* $\bar{x} \in \mathbf{S}$*. Si además* $\bar{x} \in \mathbf{S}$ *satisface* (MF)*, entonces* $T_{\mathbf{S}}(\bar{x}) = L_{\mathbf{S}}(\bar{x})$*.*

Demostración. Dividamos la demostración en partes.

1. Sea $\bar{x} \in \mathbf{S}$ y probemos que $T_{\mathbf{S}}(\bar{x}) \subseteq L_{\mathbf{S}}(\bar{x})$. Sea $d \in T_{\mathbf{S}}(\bar{x})$, luego existe $\{(t_k, d_k)\} \subseteq (0,+\infty) \times \mathbf{X}$ tal que $(t_k, d_k) \to (0, d)$ y además

   $$g_i(\bar{x} + t_k d_k) \leq 0, \quad \forall i \in \{1,\dots,p\} \quad h_j(\bar{x} + t_k d_k) = 0, \quad \forall j \in \{1,\dots,q\}$$

   Por lo tanto, dado que $g_i(\bar{x}) = 0$ para cualquier $i \in I(\bar{x})$, tenemos que

   $$\frac{g_i(\bar{x} + t_k d_k) - g_i(\bar{x})}{t_k} \leq 0, \quad \forall i \in I(\bar{x}) \quad \frac{h_j(\bar{x} + t_k d_k) - h(\bar{x})}{t_k} = 0, \quad \forall j \in \{1,\dots,q\}.$$

   Como las funciones son Fréchet diferenciables, haciendo $k \to +\infty$ obtenemos que $d \in L_{\mathbf{S}}(\bar{x})$.

2. Supongamos ahora que $\bar{x} \in \mathbf{S}$ satisface (MF) y probemos que $L_{\mathbf{S}}(\bar{x}) \subseteq T_{\mathbf{S}}(\bar{x})$. Sea $d \in L_{\mathbf{S}}(\bar{x})$ y consideremos para cada $j \in \{1,\dots,q\}$ la función $\Phi_j : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$ definida por

   $$\Phi_j(t,u) = h_j\left(\bar{x} + td + \sum_{k=1}^{q} u_k \nabla h_k(\bar{x})\right), \qquad \forall t \in \mathbb{R},\ \forall u \in \mathbb{R}^q.$$

   Denotemos por $\Phi : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}^q$ la función vectorial definida por

   $$\Phi(t,u) = (\Phi_1(t,u),\dots,\Phi_q(t,u)), \qquad \forall t \in \mathbb{R},\ \forall u \in \mathbb{R}^q.$$

Notemos que Φ es continuamente diferenciable y que $\Phi(0,0) = 0$. Más aún, tenemos que

$$\partial_{u_k}\Phi_j(0,0) = \partial_{u_j}\Phi_k(0,0) = \langle \nabla h_j(\bar{x}), \nabla h_k(\bar{x})\rangle, \quad \forall j,k \in \{1,\dots,q\}.$$

En consecuencia,

$$\nabla_u \Phi(0,0) = \begin{pmatrix} \langle \nabla h_1(\bar{x}), \nabla h_1(\bar{x})\rangle & \dots & \langle \nabla h_1(\bar{x}), \nabla h_q(\bar{x})\rangle \\ \dots & \ddots & \dots \\ \langle \nabla h_q(\bar{x}), \nabla h_1(\bar{x})\rangle & \dots & \langle \nabla h_q(\bar{x}), \nabla h_q(\bar{x})\rangle \end{pmatrix}$$

Más aún, dado que $\{\nabla h_1(\bar{x}),\dots,\nabla h_q(\bar{x})\}$ son linealmente independientes gracias a (MF), tenemos que $\nabla_u\Phi(0,0)$ es invertible, pues si $\nabla_u\Phi(0,0)u = 0$ para algún $u \in \mathbb{R}^q$, entonces

$$\left\langle \nabla h_j(\bar{x}), \sum_{k=1}^{q} u_k \nabla h_k(\bar{x}) \right\rangle = 0, \qquad \forall j \in \{1,\dots,q\}.$$

Esto a su vez implica que $\sum_{k=1}^{q} u_k \nabla h_k(\bar{x}) = 0$, y a posterior esto también implica que $u = 0$.

3. Gracias al Teorema de la Función Implícita tenemos que existen $\varepsilon > 0$ y $u : \mathbb{R} \to \mathbb{R}^q$ continuamente diferenciable en $(-\varepsilon,\varepsilon)$ tal que

$$\Phi(t,u(t)) = 0, \quad \forall t \in (-\varepsilon,\varepsilon) \quad \text{con} \quad u(0) = 0.$$

En consecuencia, la curva $x : \mathbb{R} \to \mathbf{X}$ definida por

$$x(t) = \bar{x} + t\left(d + \sum_{k=1}^{q} \frac{u_k(t)}{t}\nabla h_k(\bar{x})\right), \quad \forall t \in \mathbb{R}$$

satisface

$$h_j(x(t)) = 0, \quad \forall t \in (-\varepsilon,\varepsilon),\ \forall j \in \{1,\dots,q\}.$$

Notemos también que $\dot{u}(0) = -[\nabla_u\Phi(0,0)]^{-1}\partial_t\Phi(0,0) = 0$, pues

$$\partial_t\Phi_j(0,0) = \langle \nabla h_j(\bar{x})), d\rangle = 0, \quad \forall j \in \{1,\dots,q\}.$$

ya que $d \in L_{\mathbf{S}}(\bar{x})$. Por lo tanto, $x(0) = \bar{x}$ y $\dot{x}(0) = d$.

4. Dado que $d \in L_{\mathbf{S}}(\bar{x})$, tenemos que $\langle \nabla g_i(\bar{x}), d\rangle \le 0$ para todo $i \in I(\bar{x})$. Supongamos que la desigualdad es estricta, luego dado que las funciones son Fréchet diferenciables tenemos que

$$g_i(x(t)) = g_i(\bar{x}) + t\langle \nabla g_i(\bar{x}), d\rangle + o_i(t), \quad \forall t \in (-\varepsilon,\varepsilon),\ \forall i \in \{1,\dots,q\}, \qquad \text{con } \lim_{t\to 0}\frac{o_i(t)}{t} = 0.$$

Como $g_i(\bar{x}) = 0$ para $i \in I(\bar{x})$, vemos que podemos tomar una sucesión $\{t_k\} \subseteq (0,+\infty)$ tal que $t_k \to 0$ y $g_i(x(t_k)) \le 0$ para todo $k \in \mathbb{N}$ y por lo tanto $x(t_k) \in \mathbf{S}$. Luego para concluir basta notar que

$$x(t_k) = \bar{x} + t_k d_k, \quad \text{con} \quad d_k = d + \sum_{j=1}^{q} \frac{u_j(t_k)}{t_k}\nabla h_j(\bar{x})$$

y que $d_k \to d$, pues $\frac{u_j(t_k)}{t_k} = \frac{u_j(t_k)-u_j(0)}{t_k} \to \dot{u}_j(0) = 0$ cuando $k \to +\infty$.

5. Finalmente, si para algún índice $i \in I(\bar{x})$ tenemos que $\langle \nabla g_i(\bar{x}), d \rangle = 0$, definimos $d_\alpha = d + \alpha \bar{d}$, con $\bar{d}$ dado por (MF) y $\alpha > 0$. En este caso tenemos que $\langle \nabla g_i(\bar{x}), d_\alpha \rangle < 0$ y por lo tanto $d_\alpha \in T_{\mathbf{S}}(\bar{x})$, usando los argumentos de las partes anteriores. Finalmente, dado que $T_{\mathbf{S}}(\bar{x})$ es cerrado y $d_\alpha \to d$ si $\alpha \to 0$, concluimos que $d \in T_{\mathbf{S}}(\bar{x})$.

□

Condición de Calificación ILGA

Veremos ahora otra condición de calificación ampliamente usada y que en particular implica la condición de Mangasarian-Fromovitz. Esta es la situación cuando los gradientes de las funciones h_j y los gradientes de las restricciones activas g_i en $\bar{x}$ son linealmente independientes.

Definición 7.5. *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert. Sean $g_1, \ldots, g_p : \mathbf{X} \to \mathbb{R}$ y $h_1, \ldots, h_q : \mathbf{X} \to \mathbb{R}$ funciones Gâteaux diferenciable y sea*

$$\mathbf{S} = \left\{ x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1, \ldots, p, \quad h_j(x) = 0,\ j = 1, \ldots, q \right\}.$$

Diremos que un punto $\bar{x} \in \mathbf{S}$ satisface la condición de Independencia Lineal de Gradientes Activos *si*

(ILGA) $\quad \left\{ \nabla h_1(\bar{x}), \ldots, \nabla h_q(\bar{x}) \right\} \cup \bigcup_{i \in I(\bar{x})} \left\{ \nabla g_i(\bar{x}) \right\}$ *son linealmente independientes.*

Ahora veremos que efectivamente (ILGA) implica (MF).

Proposición 7.1. *Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert. Sean $g_1, \ldots, g_p : \mathbf{X} \to \mathbb{R}$ y $h_1, \ldots, h_q : \mathbf{X} \to \mathbb{R}$ funciones Gâteaux diferenciables, y considere el conjunto*

$$\mathbf{S} = \left\{ x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1, \ldots, p, \quad h_j(x) = 0,\ j = 1, \ldots, q \right\}.$$

Si $\bar{x} \in \mathbf{S}$ satisface (ILGA) entonces $\bar{x}$ también satisface (MF).

Demostración. Para simplificar la notación, supongamos que $I(\bar{x}) = \{1, \ldots, p\}$. Luego, basta notar que si (ILGA) se satisface, entonces el vector $(-1, \ldots, -1, 0, \ldots, 0) \in \mathbb{R}^p \times \mathbb{R}^q$ pertenece a la imagen del operador lineal continuo $A : \mathbf{X} \to \mathbb{R}^p \times \mathbb{R}^q$ definido por

$$A(d) = \left(\langle \nabla g_1(\bar{x}), d \rangle, \ldots, \langle \nabla g_p(\bar{x}), d \rangle, \langle \nabla h_1(\bar{x}), d \rangle, \ldots, \langle \nabla h_q(\bar{x}), d \rangle \right), \quad \forall d \in \mathbf{X}.$$

En efecto, si esto no fuese así, por el Teorema de Hahn-Banach (Lema 3.1), existiría un vector $(\mu, \lambda) \in \mathbb{R}^p \times \mathbb{R}^q \setminus \{0\}$ tal que

$$\left\langle \sum_{i=1}^{p} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\bar{x}), d \right\rangle < -\sum_{i=1}^{p} \mu_i, \qquad \forall d \in \mathbf{X}.$$

Tomemos $\alpha \in \mathbb{R}$ cualquiera. Evaluando en $d = \alpha \left(\sum_{i=1}^{p} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\bar{x}) \right)$, vemos que

$$\alpha \left\| \sum_{i=1}^{p} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\bar{x}) \right\|^2 < -\sum_{i=1}^{p} \mu_i.$$

Dado que $\alpha \in \mathbb{R}$ es arbitrario, concluimos que $\sum_{i=1}^{p} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\bar{x}) = 0$, luego por (ILGA) tenemos que $\mu = 0$ y $\lambda = 0$, lo que no puede ser. En particular, concluimos que existe $d \in \mathbf{X}$ tal que

$$\langle \nabla g_i(\bar{x}), d \rangle = -1,\ i = 1, \ldots, p \quad \text{y} \quad \langle \nabla h_j(\bar{x}), d \rangle = 0,\ j = 1, \ldots, q.$$

□

7.2.3. Teorema de Karush-Kuhn-Tucker

Veremos a continuación la versión general del Teorema de Kuhk-Tucker (Teorema 5.4). En este caso, y a diferencia del caso convexo, tenemos que esta condición solo serán necesaria para que un punto sea mínimo local del problema de programación matemática

(P$_{\text{PM}}$) Minimizar $f(x)$ sobre $x \in \mathbf{X}$ tales que $g_i(x) \leq 0,\ i \in \{1,\dots,p\},\ h_j(x) = 0,\ j \in \{1,\dots,q\}$.

Consideremos la función Lagrangeana asociada al problema de programación matemática (P$_{\text{PM}}$), que denotamos $L : \mathbf{X} \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R} \cup \{+\infty\}$, y que está dada por

$$L(x,\mu,\lambda) := f(x) + \sum_{i=1}^{p} \mu_i g_i(x) + \sum_{j=1}^{q} \lambda_j h_j(x), \quad \forall x \in \mathbb{R}^n,\ \mu \in \mathbb{R}^p,\ \lambda \in \mathbb{R}^q.$$

Luego el Teorema sobre condiciones de optimalidad para el problema de programación matemática es como sigue.

Teorema 7.4 (Karush-Kuhn-Tucker)**.** *Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert. Sea $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función propia, $g_1,\dots,g_p : \mathbf{X} \to \mathbb{R}$ y $h_1,\dots,h_q : \mathbf{X} \to \mathbb{R}$ funciones continuamente diferenciables. Sea $\bar{x} \in \mathbf{X}$ un mínimo local del problema de programación matemática* (P$_{\text{PM}}$)*. Supongamos que $\bar{x}$ satisface* (MF) *y que f es localmente Lipschitz continua y Gâteaux diferenciable en una vecindad de $\bar{x}$. Entonces, existen multiplicadores $\mu_1,\dots,\mu_p \geq 0$ y $\lambda_1,\dots,\lambda_q \in \mathbb{R}$ tales que*

(KKT)
$$0 = \nabla_x L(\bar{x},\mu,\lambda) = \nabla f(\bar{x}) + \sum_{i=1}^{p} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\bar{x}) \quad y \quad \mu_i g_i(\bar{x}) = 0, \quad \forall i \in \{1,\dots,p\}.$$

Demostración. Definimos el conjunto

$$\mathbf{S} = \left\{ x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1,\dots,p, \quad h_j(x) = 0,\ j = 1,\dots,q \right\}.$$

Gracias al Teorema 7.1 tenemos que (CNPO) se verifica. Además, por el Teorema 7.3 sabemos que $T_{\mathbf{S}}(\bar{x}) = L_{\mathbf{S}}(\bar{x})$, y por lo tanto para cualquier $d \in \mathbf{X}$ tenemos

$$\langle \nabla g_i(\bar{x}), d\rangle \leq 0,\ \forall i \in I(\bar{x}) \quad \text{y} \quad \left\langle \nabla h_j(\bar{x}), d \right\rangle = 0,\ \forall j \in \{1,\dots,q\} \quad \Longrightarrow \quad \langle \nabla f(\bar{x}), d\rangle \geq 0 \tag{7.1}$$

El resultado final es consecuencia entonces del Teorema de Hahn-Banach Geométrico. En efecto, consideremos el conjunto convexo cerrado y no vacío:

$$A = \left\{ v \in \mathbf{X} \,\middle|\, \exists \mu \in \mathbb{R}^p_+,\ \lambda \in \mathbb{R}^q \text{ tales que } v = \sum_{i \in I(\bar{x})} \mu_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\bar{x}) \right\}.$$

Notemos que (KKT) es equivalente a pedir $-\nabla f(\bar{x}) \in A$. Si esto no fuese cierto y dado que $\mathbf{X}$ es reflexivo (pues $\mathbf{X}$ es un espacio de Hilbert), tendríamos por Teorema de Hahn-Banach Geométrico (Lema 3.1) que existe $d \in \mathbf{X} \setminus \{0\}$ tal que

$$\sum_{i=1}^{p} \mu_i \langle \nabla g_i(\bar{x}), d\rangle + \sum_{j=1}^{q} \lambda_j \langle \nabla h_j(\bar{x}), d\rangle < -\langle \nabla f(\bar{x}), d\rangle, \qquad \forall \mu \in \mathbb{R}^p_+,\ \lambda \in \mathbb{R}^q. \tag{7.2}$$

En particular, para cualquier $i \in I(\bar{x})$, si e_i denota al i-ésimo vector canónico de $\mathbb{R}^p$, tomando $\mu = ke_i$ con $k \in \mathbb{N} \setminus \{0\}$ y $\lambda = 0$, tenemos que

$$\langle \nabla g_i(\bar{x}), d\rangle < \frac{-1}{k}\langle \nabla f(\bar{x}), d\rangle, \qquad \forall k \in \mathbb{N} \setminus \{0\}.$$

Luego haciendo $k \to +\infty$, podemos concluir que $\langle \nabla g_i(\bar{x}), d\rangle \leq 0$ para todo $i \in I(\bar{x})$. Por otro lado, tomando $\mu = 0$ y $\lambda = (\pm k, 0, \ldots, 0)$ con $k \in \mathbb{N} \setminus \{0\}$ llegamos a

$$\frac{1}{k}\langle \nabla f(\bar{x}), d\rangle < \langle \nabla h_1(\bar{x}), d\rangle < \frac{-1}{k}\langle \nabla f(\bar{x}), d\rangle, \qquad \forall k \in \mathbb{N} \setminus \{0\}.$$

Haciendo $k \to +\infty$, vemos que $\langle \nabla h_1(\bar{x}), d\rangle = 0$. Usando el mismo razonamiento para los otros índices llegamos a que $\langle \nabla h_j(\bar{x}), d\rangle = 0$ para todo $j \in \{1, \ldots, q\}$. Luego por (7.1) tenemos que $\langle \nabla f(\bar{x}), d\rangle \geq 0$, pero esto contradice (7.2) al tomar $\mu = 0$ y $\lambda = 0$. Por lo tanto, $-\nabla f(\bar{x}) \in A$ y (KKT) se verifica. □

Notemos que (KKT) se puede interpretar en términos de los puntos críticos del Lagrangiano del problema. En efecto, (KKT) es equivalente a pedir que $\bar{x}$ sea punto crítico de la función $x \mapsto L(x, \mu, \lambda)$, para algún $\mu \in \mathbb{R}^p_+$ y $\lambda \in \mathbb{R}^q$ apropiados. La heurística que hay detrás es que si $\bar{x}$ es un mínimo local del problema de programación matemática (P_{PM}), entonces es un mínimo local del problema sin restricciones

$$\text{Minimizar } L(x, \mu, \lambda) \text{ sobre todos los } x \in \mathbf{X}.$$

Esta interpretación no es del todo rigurosa, pero da una buena intuición de lo que sucede. Así mismo, la heurística descrita más arriba nos dice que para poder clasificar puntos críticos del Lagrangiano necesitamos, al igual que en el caso de optimización sin restricciones, estudiar condiciones de segundo orden que consideren segundas derivadas del Lagrangiano.

7.2.4. Condiciones de Segundo Orden

Dado que necesitamos derivadas de segundo orden, en lo que sigue de la sección asumiremos un poco más de regularidad sobre las funciones involucradas en el problema de programación matemática. En particular pediremos que las funciones sean dos veces continuamente Fréchet diferenciables.

Recuerdo : Funciones dos veces continuamente diferenciables

Una función $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ definida en un espacio de Hilbert $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ se dice *dos veces continuamente diferenciable* en $x \in \text{int}(\text{dom}(f))$ si es dos veces Fréchet diferenciable en x (en particular, continuamente diferenciable) y $\nabla^2 f(x) : \mathbf{X} \times \mathbf{X} \to \mathbb{R}$ es continuo en una vecindad de x, es decir,

$$\forall \varepsilon > 0,\ \exists r > 0 \text{ tal que } \forall y \in \mathbf{X} \quad \|x - y\| < r \implies \sup_{h,k \in \mathbb{B}_{\mathbf{X}}} |\nabla^2 f(x)(h,k) - \nabla^2 f(y)(h,k)| < \varepsilon.$$

En el caso $\mathbf{X} = \mathbb{R}^n$, esto se reduce simplemente a que las segundas derivadas parciales de f sean todas funciones continuas en una vecindad de x.

Antes de presentar las condiciones de optimalidad de segundo orden, necesitamos introducir una nueva noción de cono tangente, que es similar al cono linealizante, pero que considera solo direcciones en las que f no puede crecer.

Definición 7.6. *Sean $g_1,\dots,g_p : \mathbf{X} \to \mathbb{R}$ y $h_1,\dots,h_q : \mathbf{X} \to \mathbb{R}$ funciones Gâteaux diferenciable y sea*

$$\mathbf{S} = \left\{x \in \mathbf{X} \mid g_i(x) \leq 0,\ i = 1,\dots,p, \quad h_j(x) = 0,\ j = 1,\dots,q\right\}.$$

Definimos el cono de direcciones críticas *a* $\mathbf{S}$ *en* $x \in \mathbf{S}$ *como el conjunto*

$$K_{\mathbf{S}}(x) := \{d \in T_{\mathbf{S}}(x) \mid \langle \nabla f(x), d\rangle \leq 0\}.$$

Con esta nueva herramienta podemos ahora presentar un criterio necesario de segundo orden para que un punto sea un mínimo local del problema de programación matemática (P_{PM}).

Observación 7.2. *El siguiente resultado lo demostraremos bajo la condición de calificación* (ILGA). *El resultado sigue siendo cierto si se asume* (MF)*, sin embargo la demostración requiere herramientas de programación lineal y dualidad que no hemos estudiado en el curso.*

Teorema 7.5 (Condición Necesaria de Segundo Orden)**.** *Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert. Sea $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función propia. Sean $g_1,\dots,g_p : \mathbf{X} \to \mathbb{R}$ y $h_1,\dots,h_q : \mathbf{X} \to \mathbb{R}$ funciones dos veces continuamente diferenciables. Sea $\bar{x} \in \mathbf{X}$ un mínimo local del problema de programación matemática* (P_{PM})*. Supongamos que $\bar{x}$ satisface* (ILGA) *y que f es dos veces continuamente diferenciable en una vecindad de $\bar{x}$. Entonces, existen $\mu_1,\dots,\mu_p \geq 0$ y $\lambda_1,\dots,\lambda_q \in \mathbb{R}$ tales que* (KKT) *se satisface y que además*

$$\nabla^2_{xx} L(\bar{x},\mu,\lambda)(d,d) \geq 0, \quad \forall d \in K_{\mathbf{S}}(\bar{x}). \tag{CNSO}$$

Demostración. Dividamos la demostración en partes.

1. Supongamos que $\bar{x}$ es mínimo local que satisface (ILGA) y sea $d \in K_{\mathbf{S}}(\bar{x})$. Como $\bar{x}$ satisface (MF) (por la Proposición 7.1), gracias al Teorema 7.3 se tiene $T_{\mathbf{S}}(\bar{x}) = L_{\mathbf{S}}(\bar{x})$, de donde

   $$\langle \nabla g_i(\bar{x}), d\rangle \leq 0, \quad \forall i \in I(\bar{x}), \qquad \langle \nabla h_j(\bar{x}), d\rangle = 0, \quad \forall j \in \{1,\dots,q\}. \qquad \langle \nabla f(\bar{x}), d\rangle \leq 0.$$

   Dado que $\bar{x}$ es mínimo local, de Teorema 7.1 se deduce $\langle \nabla f(\bar{x}), d\rangle = 0$.

2. Definamos $I_d(\bar{x}) = \{i \in I(\bar{x}) \mid \langle \nabla g_i(\bar{x}), d\rangle = 0\}$ y $N_I = |I_d(\bar{x})|$. Sin perdida de generalidad asumamos que $N_I > 0$ y que $I_d(\bar{x}) = \{1,\dots,N_I\}$. Sea $\Phi\colon \mathbb{R} \times \mathbb{R}^{N_I} \times \mathbb{R}^q \to \mathbb{R}^{N_I} \times \mathbb{R}^q$ el campo vectorial cuya componentes son

   $$\Phi_i(t,\mu,\lambda) := g_i\left(\bar{x} + td + \sum_{k=1}^{N_I} \mu_k \nabla g_k(\bar{x}) + \sum_{\ell=1}^{q} \lambda_\ell \nabla h_\ell(\bar{x})\right), \quad \forall i \in \{1,\dots,N_I\}$$

   $$\Phi_j(t,\mu,\lambda) := h_j\left(\bar{x} + td + \sum_{k=1}^{N_I} \mu_k \nabla g_k(\bar{x}) + \sum_{\ell=1}^{q} \lambda_\ell \nabla h_\ell(\bar{x})\right), \quad \forall j \in \{1,\dots,q\}.$$

   Se tiene $\Phi(0,0,0) = 0$ y $\nabla_{(\mu,\lambda)}\Phi(0,0)$, la matriz Jacobiana de Φ con respecto a las variables μ y λ está dada por

   $$\begin{pmatrix} \langle \nabla g_1(\bar{x}), \nabla g_1(\bar{x})\rangle & \dots & \langle \nabla g_1(\bar{x}), \nabla g_{N_I}(\bar{x})\rangle & \langle \nabla g_1(\bar{x}), \nabla h_1(\bar{x})\rangle & \dots & \langle \nabla g_1(\bar{x}), \nabla h_q(\bar{x})\rangle \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \langle \nabla g_{N_I}(\bar{x}), \nabla g_1(\bar{x})\rangle & \dots & \langle \nabla g_{N_I}(\bar{x}), \nabla g_{N_I}(\bar{x})\rangle & \langle \nabla g_{N_I}(\bar{x}), \nabla h_1(\bar{x})\rangle & \dots & \langle \nabla g_{N_I}(\bar{x}), \nabla h_q(\bar{x})\rangle \\ \langle \nabla h_1(\bar{x}), \nabla g_1(\bar{x})\rangle & \dots & \langle \nabla h_1(\bar{x}), \nabla g_{N_I}(\bar{x})\rangle & \langle \nabla h_1(\bar{x}), \nabla h_1(\bar{x})\rangle & \dots & \langle \nabla h_1(\bar{x}), \nabla h_q(\bar{x})\rangle \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \langle \nabla h_q(\bar{x}), \nabla g_1(\bar{x})\rangle & \dots & \langle \nabla h_q(\bar{x}), \nabla g_{N_I}(\bar{x})\rangle & \langle \nabla h_q(\bar{x}), \nabla h_1(\bar{x})\rangle & \dots & \langle \nabla h_q(\bar{x}), \nabla h_q(\bar{x})\rangle \end{pmatrix}.$$

Notemos que la matriz $\nabla_{(\mu,\lambda)}\Phi(0,0)$ es invertible. En efecto, para todo $u \in \mathbb{R}^{N_I}$ y $v \in \mathbb{R}^q$ se tiene que si $\nabla_{(\mu,\lambda)}\Phi(0,0)(u,v) = 0$ entonces

$$0 = (u,v)^\top \nabla_{(\mu,\lambda)}\Phi(0,0)(u,v) = \left\| \sum_{i=1}^{N_I} u_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} v_j \nabla h_j(\bar{x}) \right\|^2.$$

Gracias a (ILGA) deducimos que $(u,v) = (0,0)$. Luego, ocupando el Teorema de la Función Implícita y dado que Φ es dos veces continuamente diferenciable, existe $\varepsilon > 0$ y funciones $\mu\colon (-\varepsilon,\varepsilon) \to \mathbb{R}^{N_I}$ y $\lambda\colon (-\varepsilon,\varepsilon) \to \mathbb{R}^q$ también dos veces continuamente diferenciables tales que $\Phi(t,\mu(t),\lambda(t)) = 0$ para todo $t \in (-\varepsilon,\varepsilon)$, con $(\mu(0),\lambda(0)) = (0,0)$.

3. Definiendo la trayectoria $x : (-\varepsilon,\varepsilon) \to \mathbf{X}$ via la fórmula

$$x(t) := \bar{x} + td + \sum_{k=1}^{N_I} \mu_k(t) \nabla g_k(\bar{x}) + \sum_{\ell=1}^{q} \lambda_\ell(t) \nabla h_\ell(\bar{x}),$$

se tiene $x(0) = \bar{x}$ y $\dot{x}(0) = d + \sum_{k=1}^{N_I} \dot{\mu}_k(0) \nabla g_k(\bar{x}) + \sum_{\ell=1}^{q} \dot{\lambda}_\ell(0) \nabla h_\ell(\bar{x})$. Además,

$$0 = \frac{d}{dt}\Phi_i(\cdot,\mu(\cdot),\lambda(\cdot))(0) = \langle \nabla g_i(\bar{x}), d \rangle + \left\langle \sum_{k=1}^{N_I} \dot{\mu}_k(0) \nabla g_k(\bar{x}) + \sum_{\ell=1}^{q} \dot{\lambda}_\ell(0) \nabla h_\ell(\bar{x}), \nabla g_i(\bar{x}) \right\rangle$$
$$0 = \frac{d}{dt}\Phi_j(\cdot,\mu(\cdot),\lambda(\cdot))(0) = \langle \nabla h_j(\bar{x}), d \rangle + \left\langle \sum_{k=1}^{N_I} \dot{\mu}_k(0) \nabla g_k(\bar{x}) + \sum_{\ell=1}^{q} \dot{\lambda}_\ell(0) \nabla h_\ell(\bar{x}), \nabla h_j(\bar{x}) \right\rangle$$

para todo $i \in \{1,\ldots,N_I\}$ y $j \in \{1,\ldots,q\}$ y se satisface $\langle \nabla g_i(\bar{x}), d \rangle = \langle \nabla h_j(\bar{x}), d \rangle = 0$. Entonces, multiplicando la primera ecuación por $\dot{\mu}_i(0)$, la segunda por $\dot{\lambda}_j(0)$ y, sumando sobre $i \in \{1,\ldots,N_I\}$ y $j \in \{1,\ldots,q\}$, se obtiene

$$0 = \left\| \sum_{k=1}^{N_I} \dot{\mu}_k(0) \nabla g_k(\bar{x}) + \sum_{\ell=1}^{q} \dot{\lambda}_\ell(0) \nabla h_\ell(\bar{x}) \right\|^2,$$

que junto con (ILGA) implican $\dot{\mu}_i(0) = \dot{\lambda}_j(0) = 0$ para todo $i \in I_d(\bar{x})$ y $j \in \{1,\ldots,q\}$. En consecuencia, deducimos que $\dot{x}(0) = d$.

4. Probemos ahora que $x(t)$ es factible para todo $t > 0$ en una vecindad de $t = 0$. En efecto, notemos que

$$0 = \Phi(t,\mu(t),\lambda(t)) = (g_1(x(t)),\ldots,g_{N_I}(x(t)),h_1(x(t)),\ldots,h_q(x(t))).$$

Si $i \notin I(\bar{x})$, entonces por continuidad de $t \mapsto g_i(x(t))$, para $t \in \mathbb{R}$ suficientemente pequeño tendríamos que $g_i(x(t)) < 0$. Por otra parte si $i \in I(\bar{x}) \setminus I_d(\bar{x})$ se tiene

$$g_i(x(t)) = g_i(\bar{x}) + \langle \nabla g_i(\bar{x}), d \rangle t + o(t) = \langle \nabla g_i(\bar{x}), d \rangle t + o(t),$$

y como $\langle \nabla g_i(\bar{x}), d \rangle < 0$ se deduce que para $t > 0$ suficientemente pequeño se tiene $g_i(x(t)) < 0$ y se tiene la factibilidad de $x(t)$ para $t > 0$ en una vecindad de $t = 0$.

5. Como $x(t)$ es factible para $t > 0$ y dos veces continuamente diferenciable, la expansión de Taylor de segundo orden de $t \mapsto f(x(t))$ en torno a $t = 0$ implica que para $t > 0$ suficientemente pequeño

$$f(\bar{x}) \leq f(x(t)) = f(\bar{x}) + \langle \nabla f(\bar{x}), d\rangle t + \frac{1}{2}\left(\nabla^2 f(\bar{x})(d,d) + \langle \nabla f(\bar{x}), \ddot{x}(0)\rangle\right) t^2 + o(t^2),$$

y como $\langle \nabla f(\bar{x}), d\rangle = 0$, dividiendo por t^2 y pasando al límite, se deduce

$$0 \leq \nabla^2 f(\bar{x})(d,d) + \langle \nabla f(\bar{x}), \ddot{x}(0)\rangle. \tag{7.3}$$

De manera similar, para $i \in \{1,\ldots,N_I\}$ y $j \in \{1,\ldots,q\}$, dado que las funciones $t \mapsto g_i(x(t))$ y $t \mapsto h_j(x(t))$ son dos veces diferenciables y nulas en el intervalo $(-\varepsilon,\varepsilon)$, estas satisfacen

$$0 = \nabla^2 g_i(\bar{x})(d,d) + \langle \nabla g_i(\bar{x}), \ddot{x}(0)\rangle, \quad \forall i \in \{1,\ldots,N_I\} \tag{7.4}$$

$$0 = \nabla^2 h_j(\bar{x})(d,d) + \left\langle \nabla h_j(\bar{x}), \ddot{x}(0)\right\rangle, \quad \forall j \in \{1,\ldots,q\}. \tag{7.5}$$

6. Sean $\mu_1,\ldots,\mu_p \geq 0$ y $\lambda_1,\ldots,\lambda_q \in \mathbb{R}$ multiplicadores asociados a $\bar{x}$ tales que (KKT) se satisface. En particular, tenemos que

$$\langle \nabla f(\bar{x}), d\rangle + \sum_{i \in I(\bar{x}) \setminus I_d(\bar{x})} \mu_i \langle \nabla g_i(\bar{x}), d\rangle + \sum_{i \in I_d(\bar{x})} \mu_i \langle \nabla g_i(\bar{x}), d\rangle + \sum_{j=1}^{q} \lambda_j \left\langle \nabla h_j(\bar{x}), d\right\rangle = 0.$$

En consecuencia $\sum_{i \in I(\bar{x}) \setminus I_d(\bar{x})} \mu_i \langle \nabla g_i(\bar{x}), d\rangle = 0$ pues todos los otros términos del lado izquierdo son cero. Ahora bien, dado que $\langle \nabla g_i(\bar{x}), d\rangle < 0$ y $\mu_i \geq 0$ para todo $i \in I(\bar{x}) \setminus I_d(\bar{x})$, concluimos que $\mu_i = 0$ cualquiera sea $i \in I(\bar{x}) \setminus I_d(\bar{x})$. Finalmente, multiplicando (7.4) por el correspondiente μ_i, (7.5) por el respectivo λ_j y sumando deducimos el resultado.

□

Ahora revisaremos una condición suficiente para que un punto que verifica las condiciones de (KKT) sea efectivamente un mínimo local del problema. Al igual que en el caso convexo, la curvatura de la función sobre el conjunto de restricciones jugará un rol importante. En este caso, esta curvatura se medirá a través de la segunda derivada del Lagrangiano.

Observación 7.3. *Es importante destacar que en el siguiente resultado ninguna condición de calificación es requerida. Sin embargo, el precio a pagar es que el espacio debe ser de dimensión finita. Existen condiciones suficiente de segundo orden en espacios de dimensión infinita, pero requieren utilizar otras nociones de cono de direcciones críticas.*

Teorema 7.6 (Condición Suficiente de Segundo Orden)**.** *Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert de dimensión finita. Sea $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función propia dos veces continuamente diferenciable en una vecindad de $\bar{x} \in int(dom(f))$. Sean $g_1,\ldots,g_p : \mathbf{X} \to \mathbb{R}$ y $h_1,\ldots,h_q : \mathbf{X} \to \mathbb{R}$ funciones dos veces continuamente diferenciables. Asumamos que*

$$\bar{x} \in \mathbf{S} = \left\{x \in \mathbf{X} \mid \ g_i(x) \leq 0,\ i = 1,\ldots,p, \quad h_j(x) = 0,\ j = 1,\ldots,q\right\}$$

y que para cada $d \in K_{\mathbf{S}}(\bar{x}) \setminus \{0\}$ existen $\mu_1,\ldots,\mu_p \geq 0$ y $\lambda_1,\ldots,\lambda_q \in \mathbb{R}$ tales que (KKT) se satisface y que además

$$\nabla^2_{xx} L(\bar{x},\mu,\lambda)(d,d) > 0. \tag{CSSO}$$

Entonces $\bar{x}$ es un mínimo local estricto del problema de programación matemática (P_{PM}).

Demostración. Supongamos por contradicción que $\bar{x}$ no es un mínimo local estricto de (P_{PM}) y, por lo tanto, existe una sucesión $\{x_k\}$ en $\mathbf{S}$ que converge a $\bar{x}$ tal que $f(x_k) \leq f(\bar{x})$. Sea $d_k = \frac{1}{\|x_k - \bar{x}\|}(x_k - \bar{x})$. Pasando a una subsucesión si es necesario, tenemos que $d_k \to d \in \mathbf{X}$ con $\|d\| = 1$ y más aún, con esto vemos que $d \in T_{\mathbf{S}}(\bar{x}) \subseteq L_{\mathbf{S}}(\bar{x})$. Por otra parte,

$$0 \geq f(x_k) - f(\bar{x}) = \langle \nabla f(\bar{x}), x_k - \bar{x} \rangle + o(\|x_k - \bar{x}\|),$$

de donde $\langle \nabla f(\bar{x}), d \rangle \leq 0$ y por lo tanto $d \in K_{\mathbf{S}}(\bar{x}) \setminus \{0\}$. Sea $(\mu, \lambda) \in \mathbb{R}^p_+ \times \mathbb{R}^q$ tales que (KKT) y (CSSO) se satisfacen para d. De (KKT) se obtiene

$$L(x_k, \mu, \lambda) = f(x_k) \leq f(\bar{x}) = L(\bar{x}, \mu, \lambda).$$

Por otras parte, dado que $\nabla_x L(\bar{x}, \mu, \lambda) = 0$, de la expansión de Taylor de orden 2 de $x \mapsto L(x, \mu, \lambda)$ en torno a $\bar{x}$ se deduce

$$0 \geq L(x_k, \mu, \lambda) - L(\bar{x}, \mu, \lambda) = \frac{1}{2} \nabla^2_{xx} L(\bar{x}, \mu, \lambda)(x_k - \bar{x}, x_k - \bar{x}) + o(\|x_k - \bar{x}\|^2)$$

y dividiendo por $\|x_k - \bar{x}\|^2$ y pasando al límite concluimos $\nabla^2_{xx} L(\bar{x}, \mu, \lambda)(d, d) \leq 0$, lo que nos lleva a una contradicción y por lo tanto $\bar{x}$ debe ser un mínimo local estricto. □

7.3. Métodos de Penalización

Ahora presentaremos algunos métodos iterativos utilizados para encontrar (o más bien aproximar) mínimos locales del problema de programación matemática. Presentaremos dos tipos de métodos, ambos basados en la idea de penalizar las restricciones y estudiar un problema auxiliar de optimización sin restricciones. El primer método que veremos es un método de penalización exterior, en el sentido que las iteraciones que generan pueden no verificar la restricción del problema original. En cambio el segundo método que veremos fuerza a que las iteraciones estén en el interior del conjunto de restricciones de desigualdad.

7.3.1. Lagrangiano Aumentado

Recordemos que el Lagrangiano (o función Lagrangiana) asociado al problema de programación matemática (P_{PM}) es la función $L : \mathbf{X} \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R} \cup \{+\infty\}$ dada por

$$L(x, \mu, \lambda) := f(x) + \sum_{i=1}^{p} \mu_i g_i(x) + \sum_{j=1}^{q} \lambda_j h_j(x), \quad \forall x \in \mathbf{X},\ \mu \in \mathbb{R}^p,\ \lambda \in \mathbb{R}^q.$$

Una propiedad interesante del Lagrangiano es que, en el caso convexo (ver Teorema 5.4), si $\bar{x}$ es una solución del problema de programación matemática, entonces es también un mínimo (global e irrestricto) de la función $x \mapsto L(x, \mu, \lambda)$ con $(\mu, \lambda) \in \mathbb{R}^p \times \mathbb{R}^q$ siendo multiplicadores asociados a $\bar{x}$. Esto sugiere que en el caso convexo, que si conociésemos los multiplicadores, minimizar sin restricciones la función $x \mapsto L(x, \mu, \lambda)$ sería equivalente a resolver el problema de programación matemática. Desafortunadamente, fuera del caso convexo esto no es cierto y un mínimo local del problema de programación matemática no es necesariamente un mínimo local del Lagrangiano.

Ejemplo 7.3.1. *Considere el problema*

$$\textit{Minimizar } 1-x-\frac{1}{3}x^3 \textit{ sobre los } x\in\mathbb{R} \textit{ tales que } x\leq 0.$$

No es difícil ver que $\bar{x}=0$ es el mínimo (global) del problema. Además, imponiendo las condiciones de (KKT) se tiene que el multiplicador asociado a la restricción es $\mu=1$. Sin embargo, la función

$$x\mapsto L(x,1)=1-\frac{1}{3}x^3$$

es no acotada y $\bar{x}=0$ es sólo un punto crítico de $x\mapsto L(x,1)$ pero no es un mínimo local.

Para evitar la clase de problemas descritos con el ejemplo anterior se introduce una función llamada *Lagrangiano aumentado* del problema de programación matemática. En adelante, para simplificar la exposición, nos enfocaremos en el caso de restricciones de igualdad, es decir, en el problema

$$(\mathrm{P_I}) \qquad \text{Minimizar } f(x) \text{ sobre los } x\in\mathbf{X} \text{ tales que } h_j(x)=0,\ j\in\{1,\dots,q\}.$$

Observación 7.4. *Para el caso con restricciones de desigualdad usualmente se agrega una variable adicional (llamada holgura) y se considera el problema de optimización equivalente:*

$$\textit{Minimizar } f(x) \textit{ sobre } (x,y)\in\mathbf{X}\times\mathbb{R}^p \textit{ tales que } g_i(x)+y_i^2=0,\ i\in\{1,\dots,p\},\ h_j(x)=0,\ j\in\{1,\dots,q\}.$$

Dado $r>0$, el Lagrangiano aumentado del problema ($\mathrm{P_I}$) es la función $L_r:\mathbf{X}\times\mathbb{R}^q\to\mathbb{R}\cup\{+\infty\}$ dada por

$$L_r(x,\lambda):=f(x)+\sum_{j=1}^{q}\lambda_j h_j(x)+\frac{r}{2}\sum_{j=1}^{q}h_j^2(x),\quad \forall x\in\mathbf{X},\ \lambda\in\mathbb{R}^q.$$

Ejemplo 7.3.2. *Notemos que en el Ejemplo 7.3.1 el Lagrangiano aumentado (transformado la restricción de desigualdad por igualdad agregando la variable de holgura) es*

$$L_r(x,y,\lambda)=1-x-\frac{1}{3}x^3+\lambda(x+y^2)+\frac{r}{2}(x+y^2)^2.$$

Imponiendo las condiciones de (KKT) se tiene que el multiplicador asociado a la restricción es $\lambda=1$ y que necesariamente $\bar{y}=0$. Por lo tanto

$$L_r(x,y,1)=1-\frac{1}{3}x^3+y^2+\frac{r}{2}(x+y^2)^2,\qquad \forall x,y\in\mathbb{R}.$$

No es difícil ver, usando (CSSO) para problemas irrestrictos, que $(\bar{x},\bar{y})=(0,0)$ es efectivamente un mínimo local (estricto) de $(x,y)\mapsto L_r(x,y,1)$ pues la matriz Hessiana en $(\bar{x},\bar{y})=(0,0)$ es la matriz diagonal cuyas entradas son r y 2.

La característica descrita en el ejemplo anterior es justamente la principal motivación de introducir el Lagrangiano aumentado.

Teorema 7.7. *Sea $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ un espacio de Hilbert de dimensión finita. Sea $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ una función propia dos veces continuamente diferenciable en una vecindad de $\bar{x} \in int(dom(f))$. Sean $h_1, \ldots, h_q : \mathbf{X} \to \mathbb{R}$ funciones dos veces continuamente diferenciables. Asumamos que $\bar{x}$ es un mínimo local de (P_I), tal que (KKT) se cumple para algún $\lambda \in \mathbb{R}^q$ y tal que*

$$\nabla^2_{xx} L(\bar{x}, \lambda)(d,d) > 0, \qquad \forall d \in \mathbf{X} \setminus \{0\} \text{ tal que } \langle \nabla h_j(\bar{x}), d \rangle = 0, \ \forall j = 1, \ldots, q.$$

Entonces existe $r_0 \in \mathbb{R}$ tal que para todo $r \geq r_0$ tenemos que $\bar{x}$ es un mínimo local estricto del Lagrangiano aumentado $L_r(\cdot, \lambda)$ del problema de programación matemática (P_I).

Demostración. Notemos que como $h_j(\bar{x}) = 0$ para todo $j \in \{1, \ldots, q\}$, se tiene

$$\nabla_x L_r(\bar{x}, \lambda) = \nabla f(\bar{x}) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\bar{x}) + r \sum_{j=1}^{q} h_j(\bar{x}) \nabla h_j(\bar{x}) = \nabla L(\bar{x}, \lambda) = 0,$$

por lo que basta demostrar, por CSSO en el caso irrestricto, que existe r suficientemente grande tal que el operador bilineal

$$\begin{aligned} \nabla^2_{xx} L_r(\bar{x}, \lambda) &= \nabla^2 f(\bar{x}) + \sum_{j=1}^{q} \lambda_j \nabla^2 h_j(\bar{x}) + r \sum_{j=1}^{q} \nabla h_j(\bar{x}) \nabla h_j(\bar{x})^\top + r \sum_{j=1}^{q} h_j(\bar{x}) \nabla^2 h_j(\bar{x}) \\ &= \nabla^2_{xx} L(\bar{x}, \lambda) + r \sum_{j=1}^{q} \nabla h_j(\bar{x}) \nabla h_j(\bar{x})^\top \end{aligned}$$

es definido positivo. Por contradicción, supongamos que existe una sucesión $r_k \to \infty$ y $d_k \in \mathbf{X}$ tales que

$$\nabla^2_{xx} L_r(\bar{x}, \lambda)(d_k, d_k) = \nabla^2_{xx} L(\bar{x}, \lambda)(d_k, d_k) + r_k \sum_{j=1}^{q} |\langle \nabla h_j(\bar{x}), d_k \rangle|^2 \leq 0. \tag{7.6}$$

Dividiendo (7.6) por $\|d_k\|^2$, podemos asumir que $\|d_k\| = 1$ en la desigualdad anterior y, tomando una subsucesión si fuese necesario, podemos asumir que $d_k \to d \neq 0$. Por otra parte, si dividimos (7.6) por r_k y usamos que $\nabla^2_{xx} L(\bar{x}, \lambda)(d_k, d_k)$ es acotada, pasando al límite se obtiene $\langle \nabla h_j(\bar{x}), d \rangle = 0$ para todo $j \in \{1, \ldots, q\}$. Finalmente, como (7.6) implica que $\nabla^2_{xx} L(\bar{x}, \lambda)(d_k, d_k) \leq 0$ para todo $k \in \mathbb{N}$, llegamos a una contradicción pues hemos demostrado que $\nabla^2_{xx} L(\bar{x}, \lambda)(d,d) \leq 0$, con $\|d\| = 1$. □

Esquema Algorítmico

La noción de Lagrangiano aumentado puede ser usado para construir algoritmo. Dado que a priori uno no tiene información sobre el multiplicador asociado a un mínimo local, la búsqueda que debe realizar un algoritmo basado en el Lagrangiano aumentado debe actualizar tanto la variable x como la variable del multiplicador λ. Notemos que si $\lambda \in \mathbb{R}^p$ y $r > 0$ fuesen dados, y $\bar{x} \in \mathbf{X}$ fuese un mínimo local del Lagrangiano aumentado entonces tendríamos que

$$\nabla_x L_r(\bar{x}, \lambda) = \nabla f(\bar{x}) + \sum_{j=1}^{p} \left(\lambda_j + r h_j(\bar{x})\right) \nabla h_j(\bar{x}) = 0.$$

Luego, para que $\bar{x}$ tenga opciones de ser un mínimo local de (P_I) debería verificar

$$h_j(\bar{x}) = 0 \quad \text{y} \quad \lambda_j + rh_j(\bar{x}) = \lambda_j, \qquad \forall j \in \{1,\dots,q\}.$$

El siguiente método iterativo, que presentamos sólo a modo de información, sin discusión sobre su convergencia, utiliza las ideas descritas más arriba. Cabe mencionar que este algoritmo se espera que converja tomando en cada iteración r más grande, de forma de forzar que λ converja a algún multiplicador que verifique (KKT).

	MÉTODO DE LOS MULTIPLICADORES
1.	Tomar $\lambda \in \mathbb{R}$ y $r > 0$.
2.	Calcular $x \in \arg\min_{\mathbf{X}}(L_r(\cdot,\lambda))$.
3.	Si x satisface $h_j(x) \simeq 0$ para todo $j \in \{1,\dots,q\}$ parar.
3.	Definir $\beta_j = \lambda_j + rh_j(\bar{x})$ para cada $j \in \{1,\dots,q\}$.
4.	Actualizar $\lambda = \beta$ y $r > 0$ (de ser necesario), y volver al paso 2.

7.3.2. Barrera Logarítmica

Notemos que el método del Lagrangiano Aumentado permite generar una secuencia de puntos que no satisfacen las restricciones. En este sentido, el algoritmo se considera ser un método de *punto exterior*. Ahora veremos un método que fuerza a las iteraciones a estar en el interior del conjunto de restricciones de desigualdad penalizando el acercarse a la frontera. Esta clase de algoritmos se conoce como método de *punto interior*. Por simplicidad nos enfocaremos en el caso con sólo restricciones de desigualdad, es decir,

$$(P_D) \qquad \text{Mininimzar } f(x) \text{ sobre los } x \in \mathbf{X} \text{ tales que } g_i(x) \leq 0,\ i \in \{1,\dots,p\}$$

Para estudiar mínimos locales del problema (P_D) se propone estudiar, para $\varepsilon > 0$ dado, los mínimos locales de la aproximación de barrera logarítmica $f_\varepsilon : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ definida por

$$f_\varepsilon : x \mapsto \begin{cases} f(x) - \varepsilon \sum_{i=1}^{p} \log(-g_i(x)) & g_i(x) < 0,\ \forall i \in \{1,\dots,p\}, \\ +\infty & \text{si no,} \end{cases}.$$

La idea del método consiste encontrar un mínimo de f_ε, denotado por lo general por $x(\varepsilon)$ y luego estudiar el comportamiento de $\varepsilon \mapsto x(\varepsilon)$ hacia algún mínimo local de (P_D) cuando $\varepsilon \to 0$. Notar que, por la forma de la aproximación de barrera logarítmica, tenemos que

$$g_i(x(\varepsilon)) < 0,\ \forall i \in \{1,\dots,p\},\ \forall \varepsilon > 0.$$

Más aún, usando la (CNPO) sobre f_ε se tiene que

$$(KKT_\varepsilon) \qquad \nabla f_\varepsilon(x(\varepsilon)) = \nabla f(x(\varepsilon)) + \sum_{i=1}^{p} \mu_i(\varepsilon) \nabla g_i(x(\varepsilon)) = 0, \quad \text{donde } \mu(\varepsilon) := -\frac{\varepsilon}{g_i(x(\varepsilon))} > 0.$$

En este caso los $\mu_i(\varepsilon)$ juegan el rol de multiplicadores aproximados y en consecuencia se espera que el límite de $\mu_i(\varepsilon)$ cuando $\varepsilon \to 0$ sea un multiplicador asociado a un mínimo local de (P_D).

Proposición 7.2. *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert de dimensión finita. Sea* $f : \mathbf{X} \to \mathbb{R}$ *una función continua. Sean* $g_1, \ldots, g_p : \mathbf{X} \to \mathbb{R}$ *funciones continua. Asumamos que*

$$\mathbf{S} = \{x \in \mathbf{X} \mid \ g_i(x) \leq 0,\ i \in \{1, \ldots, p\}\}$$

es acotado y su interior es denso en $\mathbf{S}$*. Entonces, para todo* $\varepsilon > 0$ *existe* $x(\varepsilon) \in \arg\min_{\mathbf{X}}(f_\varepsilon)$*. Más aún, todo punto de acumulación de* $\{x(\varepsilon) : \ \varepsilon > 0\}$ *es solución de* ($\mathrm{P_D}$)*, esto es, toda sucesión convergente de la forma* $\{x(\varepsilon_k)\}$ *converge a un mínimo del problema* ($\mathrm{P_D}$)*, donde* $\varepsilon_k \to 0^+$ *cuando* $k \to +\infty$*.*

Demostración. Dividamos la demostración en dos partes. Primero veamos la existencia de un mínimo y luego estudiemos la convergencia de la trayectoria.

1. Por composición de funciones, no es difícil ver que f_ε es continua en

$$\operatorname{int}(\mathbf{S}) = \{x \in \mathbf{X} \mid \ g_i(x) < 0,\ i \in \{1, \ldots, p\}\}.$$

   Notemos también que $f_\varepsilon = +\infty$ si $x \notin \mathbf{S}$. Más aún, para cualquier sucesión $\{x_k\} \subseteq \operatorname{int}(\mathbf{S})$ se tiene que si $g_i(x_k) \to 0$ para algún $i \in \{1, \ldots, p\}$ entonces $f_\varepsilon(x_k) \to +\infty$. Por lo tanto, tenemos que f_ε es semicontinua inferior. Por otro lado, f_ε es propia pues $\operatorname{int}(\mathbf{S}) \neq \emptyset$. Finalmente, para determinar la existencia a través del Teorema de Wierestrass-Hilbert-Tonelli (Teorema 2.1) nos bastará ver que los conjuntos de subnivel de f_ε son acotados. Pero esto es una consecuencia directa del hecho que $\operatorname{dom}(f_\varepsilon) \subseteq \operatorname{int}(\mathbf{S})$ y del hecho que $\mathbf{S}$ es acotado. Luego la existencia de $x(\varepsilon)$ para cualquier $\varepsilon > 0$ está garantizada.

2. Estudiemos ahora los puntos de acumulación de la trayectoria $\varepsilon \mapsto x(\varepsilon)$ cuando $\varepsilon \to 0$. Sea $\{\varepsilon_k\} \subseteq (0, +\infty)$ tal que $\varepsilon_k \to 0$ cuando $k \to +\infty$. Supongamos que $x_k := x(\varepsilon_k)$ converge a un cierto $\bar{x} \in \mathbb{R}^n$. Dado que $x_k \in \operatorname{int}(\mathbf{S})$ y $\mathbf{S}$ es cerrado, tenemos que $\bar{x} \in \mathbf{S}$. Ahora bien, para cualquier $k \in \mathbb{N}$, por definición de x_k tenemos

$$f_{\varepsilon_k}(x_k) = f(x_k) - \varepsilon_k \sum_{i=1}^{p} \log(-g_i(x_k)) \leq f(x) - \varepsilon_k \sum_{i=1}^{p} \log(-g_i(x)), \qquad \forall x \in \operatorname{int}(\mathbf{S}). \tag{7.7}$$

   Por otro lado, para cada $i \in I(\bar{x})$, y por continuidad de g_i, tenemos que $g_i(x_k) \geq -1$ para todo $k \in \mathbb{N}$ suficientemente grande. En consecuencia, dado que $I(\bar{x})$ es finito, $\exists k_0 \in \mathbb{N}$ tal que

$$\log(-g_i(x_k)) \leq 0 \quad \forall i \in I(\bar{x}),\ \forall k \ \geq k_0.$$

   Notemos también que si $i \notin I(\bar{x})$, entonces la sucesión $\{\log(-g_i(x_k))\}$ permanece acotada y por lo tanto

$$\varepsilon_k \log(-g_i(x_k)) \to 0 \quad \text{ si } k \to +\infty.$$

   Finalmente, de (7.7) obtenemos que para $k \in \mathbb{N}$ suficientemente grande

$$f(x_k) - \varepsilon_k \sum_{i \notin I(\bar{x})} \log(-g_i(x_k)) \leq f(x) - \varepsilon_k \sum_{i=1}^{p} \log(-g_i(x)), \qquad \forall k \in \mathbb{N},\ \forall x \in \operatorname{int}(\mathbf{S}).$$

   Luego, pasando al límite vemos que $f(\bar{x}) \leq f(x)$ para todo $x \in \operatorname{int}(\mathbf{S})$. Finalmente, dado que $\bar{x} \in \mathbf{S}$ y $\operatorname{int}(\mathbf{S})$ es denso en $\mathbf{S}$, usando la continuidad de f concluimos que $\bar{x} \in \operatorname{sol}(\mathrm{P_D})$.

□

Observación 7.5. *El hecho que int*$(\mathbf{S})$ *sea denso en* $\mathbf{S}$ *es importante, pues de no ser así la convergencia a un mínimo del problema* (P_D) *no puede ser asegurada.*

El resultado anterior muestra que la trayectoria $\varepsilon \mapsto x(\varepsilon)$ se acumula en torno al conjunto de mínimos de f_ε cuando $\varepsilon \to 0^+$. Es importante destacar que en el resultado anterior, la existencia y convergencia de la trayectoria $\varepsilon \mapsto x(\varepsilon)$ está fuertemente ligada a que el conjunto factible es compacto. Ahora veremos un resultado un poco más general que no requiere esas hipótesis y que muestra la convergencia a un mínimo local estricto de (P_D). El resultado también provee la convergencia de los multiplicadores aproximados asociados a la trayectoria.

Teorema 7.8. *Sea* $(\mathbf{X}, \langle\cdot,\cdot\rangle)$ *un espacio de Hilbert de dimensión finita. Sea* $f : \mathbf{X} \to \mathbb{R} \cup \{+\infty\}$ *una función propia dos veces continuamente diferenciable en una vecindad de* $\bar{x} \in int(dom(f))$. *Sean* $g_1, \ldots, g_p : \mathbf{X} \to \mathbb{R}$ *funciones dos veces continuamente diferenciables tal que la condición de calificación* (ILGA) *se verifica en* $\bar{x}$. *Asumamos que* $\bar{x}$ *es un mínimo local de* (P_D) *que verifica* (KKT) *para algún* $\bar{\mu} \in \mathbb{R}^p$ *con* complementaridad estricta, *es decir,* $\bar{\mu}_i > 0$ *para todo* $i \in I(\bar{x})$, *además de la condición suficiente de segundo orden*

$$\nabla^2_{xx} L(\bar{x}, \bar{\mu})(d,d) > 0, \qquad \forall d \in \mathbf{X} \setminus \{0\} \text{ tal que } \langle \nabla g_i(\bar{x}), d\rangle = 0,\ \forall i \in I(\bar{x}).$$

Entonces existe una única trayectoria $\varepsilon \mapsto (x(\varepsilon), \mu(\varepsilon))$ *continuamente diferenciable en una vecindad de* $\varepsilon = 0$ *que verifica* (KKT_ε) *tal que* $x(0) = \bar{x}$ *y* $\mu(0) = \bar{\mu}$. *Más aún, para cada* $\varepsilon > 0$ *suficientemente pequeño se tiene que* $x(\varepsilon)$ *es un mínimo local estricto de* f_ε.

Demostración. Para simplificar la notación, consideremos el caso $\mathbf{X} = \mathbb{R}^n$. El caso $\mathbf{X}$ general se obtiene de usar la isometría canónica entre $\mathbb{R}^n$ y un espacio de Hilbert de dimensión finita.

Dado $i \in \{1, \ldots, p\}$, definamos las funciones $F_i : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^p$ definidas por

$$F_i(\varepsilon, x, \mu) = \mu_i g_i(x) + \varepsilon, \qquad \forall (\varepsilon, x, \mu) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p.$$

Consideremos además los campos vectoriales $F : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ y $G : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ dados por

$$F(\varepsilon, x, \mu) = (F_1(\varepsilon, x, \mu), \ldots, F_p(\varepsilon, x, \mu)) \text{ y } G(x, \mu) = \nabla f(x) + \sum_{i=1}^m \mu_i \nabla g_i(x), \quad \forall (\varepsilon, x, \mu) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p.$$

Por construcción, ambos campos vectoriales son continuamente diferenciables. Dado que $\bar{x}$ es un mínimo local de (P_D) que verifica (KKT) para algún $\bar{\mu} \in \mathbb{R}^p$, tenemos que $(0, \bar{x}, \bar{\mu})$ es solución de la ecuación

$$\Phi(\varepsilon, x, \mu) = 0, \quad \text{donde } \Phi(\varepsilon, x, \mu) := (F(\varepsilon, x, \mu), G(x, \mu)).$$

Luego, la existencia de una única trayectoria $\varepsilon \mapsto (x(\varepsilon), \mu(\varepsilon))$ continuamente diferenciable en una vecindad de $\varepsilon = 0$ que verifica (KKT_ε) tal que $x(0) = \bar{x}$ y $\mu(0) = \bar{\mu}$ es una consecuencia del Teorema de la Función Implícita. En efecto, notemos que

$$\nabla_{(x,\mu)} \Phi(\varepsilon, x, \mu) = \begin{bmatrix} \nabla_x F(\varepsilon, x, \mu) & \nabla_\mu F(\varepsilon, x, \mu) \\ \nabla_x G(x, \mu) & \nabla_\mu G(x, \mu) \end{bmatrix}, \quad \forall (\varepsilon, x, \mu) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p.$$

Sigue que

$$\nabla_{(x,\mu)}\Phi(0,\bar{x},\bar{\mu})(d,\nu) = \begin{pmatrix} \bar{\mu}_1 \nabla g_1(\bar{x})^\top d + \nu_1 g_1(\bar{x}) \\ \vdots \\ \bar{\mu}_p \nabla g_p(\bar{x})^\top d + \nu_p g_p(\bar{x}) \\ \nabla^2_{xx} L(\bar{x},\bar{\mu}) d + \sum_{i=1}^{p} \nu_i \nabla g_i(\bar{x}) \end{pmatrix}, \qquad \forall (d,\nu) \in \mathbb{R}^n \times \mathbb{R}^p.$$

En particular, si $\nabla_{(x,\mu)}\Phi(0,\bar{x},\bar{\mu})(d,\nu) = 0$ para ciertos $(d,\nu) \in \mathbb{R}^n \times \mathbb{R}^p$, por complementaridad estricta, para cada $i \in I(\bar{x})$ tenemos que $\nabla g_i(\bar{x})^\top d = 0$ y $\nu_i = 0$ para cada $i \notin I(\bar{x})$. Notemos además que si $d \neq 0$, entonces multiplicando por d la última ecuación tendríamos que

$$0 = d^\top \nabla^2_{xx} L(\bar{x},\bar{\mu}) d + \sum_{i \in I(\bar{x})} \nu_i \nabla g_i(\bar{x})^\top d = d^\top \nabla^2_{xx} L(\bar{x},\bar{\mu}) d.$$

Sin embargo esto contradice la condición suficiente de segundo orden del enunciado. Por lo tanto $d = 0$. Esto a su vez implica que

$$0 = \nabla^2_{xx} L(\bar{x},\bar{\mu}) d + \sum_{i=1}^{p} \nu_i \nabla g_i(\bar{x}) = \sum_{i \in I(\bar{x})} \nu_i \nabla g_i(\bar{x}).$$

Luego, por (ILGA) tenemos que $\nu_i = 0$ para cada $i \in I(\bar{x})$, y en consecuencia $\nu = 0$ y por lo tanto la matriz $\nabla_{(x,\mu)}\Phi(0,\bar{x},\bar{\mu})$ es invertible. Gracias al Teorema de la Función Implícita, existe una única trayectoria $\varepsilon \mapsto (x(\varepsilon),\mu(\varepsilon))$ continuamente diferenciable en una vecindad de $\varepsilon = 0$ que verifica (KKT_ε) tal que $x(0) = \bar{x}$ y $\mu(0) = \bar{\mu}$.

Resta ver que $x(\varepsilon)$ es un mínimo local estricto de f_ε. Para esto bastará estudiar la segunda derivada de la función f_ε y luego aplicar la (CSSO). Notemos que

$$\nabla^2 f_\varepsilon(x) = \nabla^2 f(x) + \sum_{i=1}^{p} \left(\frac{\varepsilon}{g_i(x)^2} \nabla g_i(x) \nabla g_i(x)^\top - \frac{\varepsilon}{g_i(x)} \nabla^2 g_i(x) \right), \quad \forall x \in \operatorname{dom}(f_\varepsilon).$$

Por lo tanto, evaluando en $x = x(\varepsilon)$ tenemos

$$\nabla^2 f_\varepsilon(x(\varepsilon)) = \nabla^2_{xx} L(x(\varepsilon),\mu(\varepsilon)) + \sum_{i=1}^{p} \frac{\mu_i(\varepsilon)^2}{\varepsilon} \nabla g_i(x(\varepsilon)) \nabla g_i(x(\varepsilon))^\top.$$

Sea $d \in \mathbb{R}^n \setminus \{0\}$. Separemos el resto de la demostración en dos casos:

1. Supongamos $\nabla g_i(\bar{x})^\top d = 0$ para cualquier $i \in I(\bar{x})$. Usando la condición suficiente de segundo orden tenemos que
$$d^\top \nabla^2_{xx} L(\bar{x},\bar{\mu}) d > 0.$$
Por lo tanto, por continuidad, para $\varepsilon > 0$ suficientemente pequeño tendremos
$$d^\top \nabla^2_{xx} L(x(\varepsilon),\mu(\varepsilon)) d > 0.$$
Sigue que
$$d^\top \nabla^2 f_\varepsilon(x(\varepsilon)) d = d^\top \nabla^2_{xx} L(x(\varepsilon),\mu(\varepsilon)) d + \sum_{i=1}^{p} \frac{\mu_i(\varepsilon)^2}{\varepsilon} (\nabla g_i(x(\varepsilon))^\top d)^2 \geq d^\top \nabla^2_{xx} L(x(\varepsilon),\mu(\varepsilon)) d > 0$$

2. Supongamos ahora que $\nabla g_i(\bar{x})^\top d \neq 0$ para algún $i \in I(\bar{x})$. No es difícil ver que

$$d^\top \nabla^2 f_\varepsilon(x(\varepsilon))d \geq d^\top \nabla^2_{xx} L(x(\varepsilon),\mu(\varepsilon))d + \sum_{i=1}^{p} \frac{\mu_i(\varepsilon)^2}{\varepsilon}(\nabla g_i(x(\varepsilon))^\top d)^2.$$

   Por otro lado, cuando $\varepsilon \to 0$ tenemos que

$$\nabla^2_{xx} L(x(\varepsilon),\mu(\varepsilon)) \to \nabla^2_{xx} L(\bar{x},\bar{\mu}), \quad \nabla g_i(x(\varepsilon))^\top d \to \nabla g_i(\bar{x})^\top d \neq 0 \text{ y } \mu_i(\varepsilon) \to \bar{\mu}_i > 0.$$

   En particular tenemos que $d^\top \nabla^2 f_\varepsilon(x(\varepsilon))d \to +\infty$ si $\varepsilon \to 0$. Por lo tanto, $d^\top \nabla^2 f_\varepsilon(x(\varepsilon))d > 0$ para $\varepsilon > 0$ suficientemente pequeño.

Finalmente, dado que $d^\top \nabla^2 f_\varepsilon(x)d > 0$ para cualquier $d \in \mathbb{R}^n \setminus \{0\}$ y $\varepsilon > 0$ pequeño, por Teorema 6.3, tenemos que $x(\varepsilon)$ es un mínimo local estricto de f_ε para $\varepsilon > 0$ suficientemente pequeño. □

Esquema Algorítmico

Al igual que en la parte anterior describiremos el esquema general que tiene un algoritmo basado en la aproximación de barrera logarítmica. La idea esencial del método es que en cada iteración se resuelve un sub problema de optimización sin restricciones para luego actualizar el parámetro de penalización. La convergencia del método estará entonces dada por el hecho que $\varepsilon \mapsto x(\varepsilon)$ converge a un mínimo local del problema original si $\varepsilon \to 0^+$.

	MÉTODO DE PENALIZACIÓN
1.	Tomar $\varepsilon > 0$, $\tau \in (0,1)$ y $x_0 \in \mathbf{X}$.
2.	Calcular $x \in \arg\min_{\mathbf{X}}(f_\varepsilon)$.
3.	Si $\|x - x_0\| \simeq 0$ parar.
4.	Actualizar $x_0 = x$, $\varepsilon \leftarrow \varepsilon\tau$, y volver al paso 2.

7.4. Ejercicios

1. CARACTERIZACIONES DEL CONO TANGENTE

   Sea $\mathbf{S} \subseteq \mathbb{R}^n$ un conjunto dado. Demuestre que

   $$T_{\mathbf{S}}(x) = \left\{ d \in \mathbb{R}^n \,\middle|\, \liminf_{t \to 0^+} \frac{\operatorname{dist}(x+td, \mathbf{S})}{t} \leq 0 \right\}, \qquad \forall x \in \mathbf{S}.$$

2. CONO NORMAL Y CONO TANGENTE

   Sea $(\mathbf{X}, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert y $\mathbf{S} \subseteq \mathbf{X}$ convexo no vacío. Demuestre que

   $$\eta \in N_{\mathbf{S}}(x) \quad \Longleftrightarrow \quad \langle \eta, d \rangle \leq 0, \quad \forall d \in T_{\mathbf{S}}(x),$$

3. CONDICIÓN SUFICIENTE DE PRIMER ORDER

   Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función continua y Gâteaux diferenciable en una vecindad de $\bar{x} \in \mathbf{S}$. Supongamos que

   $$\nabla f(\bar{x})^\top d > 0, \quad \forall d \in T_{\mathbf{S}}(\bar{x}) \setminus \{0\}.$$

   Pruebe que $\bar{x}$ es un mínimo local estricto del problema general de Optimización No Lineal (P).

4. MULTIPLICADORES DE KKT

   Sea $\bar{x} \in \mathbf{S} := \{x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, ..., m,\ h_j(x) = 0, j = 1, ..., p\}$ tal que las funciones g_i y h_j son diferenciables en $\bar{x}$, $\forall i = 1, ..., m,\ \forall j = 1, ..., p$. $\bar{x} \in \mathbf{S}$

   a) Demostrar que $\bar{x}$ satisface la calificación de restricciones de Mangasarian-Fromovitz (MF) ssi:

   $$\sum_{j=1}^{p} \lambda_j \nabla h_j(\bar{x}) + \sum_{i \in I_0(\bar{x})} \mu_i \nabla g_i(\bar{x}) = 0 \text{ con } \mu_i \geq 0 \quad \Longrightarrow \quad \lambda_j = \mu_i = 0, \forall j = 1, .., p, i \in I_0(\bar{x}).$$

   b) Para el problema mín$\{f(x) : x \in C\}$, supongamos que el conjunto $\Lambda(\bar{x})$ de multiplicadores de Lagrange asociados a $\bar{x}$ es no vacío. Pruebe $\bar{x}$ satisface (MF) ssi el conjunto $\Lambda(\bar{x})$ es acotado.