1

Sea $(X, \|\cdot\|)$ un espacio real normado, $f : X \to \mathbb{R}$ un funcional convexo y $x_0 \in X$.

(a) Pruebe que $\partial f(x_0)$ es un conjunto convexo (contenido en el conjunto de funcionales lineales continuos $l : X \to \mathbb{R}$).

(b) Sea $x^* : X \to \mathbb{R}$ un funcional lineal y continuo. Pruebe que

(b.1) $\partial(f + x^*)(x_0) = x^* + \partial f(x_0)$,

(b.2) $\partial(\lambda f)(x_0) = \lambda \partial f(x_0)$.

(a)

Para probar (a) primero vemos que por el Teorema 3.26 del texto base como f es convexo el conjunto $\partial f(x_0)$ es no vació. Recordamos que la definición de $\partial f(x_0)$ es,

$$\partial f(x_0) = \{l \in X^* | f(x) \geq f(x_0) + l(x - x_0),\ \forall x \in X\},$$

donde X^* es el espacio dual de X.

Para todo $l_1, l_2 \in \partial f(x_0)$ y para todo $\lambda \in [0, 1]$, dado $x \in X$ tenemos,

$$(\lambda l_1 + (1 - \lambda)l_2)(x - x_0) = \lambda l_1(x - x_0) + (1 - \lambda)l_2(x - x_0).$$

Por tanto,

$$\begin{aligned} f(x_0) + (\lambda l_1 + (1 - \lambda)l_2)(x - x_0) &= f(x_0) + \lambda l_1(x - x_0) + (1 - \lambda)l_2(x - x_0) \\ &= \lambda(f(x_0) + l_1(x - x_0)) + (1 - \lambda)(f(x_0) + l_2(x - x_0)) \\ &\leq \lambda f(x) + (1 - \lambda)f(x) \\ &= f(x). \end{aligned}$$

Y por tanto $(\lambda l_1 + (1 - \lambda)l_2) \in \partial f(x_0)$, lo que implica que $\partial f(x_0)$ es convexo.

(b.1)

Vemos que $\partial(f + x^*)(x_0) \subset x^* + \partial f(x_0)$. Para todo $l \in \partial(f + x^*)(x_0)$ tenemos que

$$(f+x^*)(x) \geq (f+x^*)(x_0)+l(x-x_0) \Rightarrow f(x_0) \geq f(x_0)+l(x-x_0)-x^*(x-x_0) = f(x_0)+(l-x^*)(x-x_0),$$

dado que x^* es un funcional lineal y continuo, concluimos que $l - x^* \in \partial f(x_0)$ y esto implica que $l \in x^* + \partial f(x_0)$.

Veamos ahora que $x^* + \partial f(x_0) \subset \partial(f + x^*)(x_0)$. Sea $l \in \partial f(x_0)$ entonces,

$$(f + x^*)(x_0) + (l + x^*)(x - x_0) = f(x_0) + l(x - x_0) + x^*(x_0) + x^*(x - x_0) \leq f(x) + x^*(x),$$

por tanto $l + x^* \in \partial(f + x^*)(x_0)$, que implica $x^* + \partial f(x_0) \subset \partial(f + x^*)(x_0)$ y consecuentemente $x^* + \partial f(x_0) = \partial(f + x^*)(x_0)$.

(b.2)

Igual que en el apartado anterior comprobamos ambas inclusiones. Veamos que $\partial(\lambda f)(x_0) \subset \lambda\partial f(x_0)$. Para todo $l \in \partial(\lambda f)(x_0)$, tenemos

$$(\lambda f)(x) \geq (\lambda f)(x_0) + l(x - x_0) \Rightarrow f(x) \geq f(x_0) + \frac{1}{\lambda} l(x - x_0),$$

por tanto $\frac{1}{\lambda} l \in \partial f(x_0)$ que implica $l \in \lambda\partial f(x_0)$.

Veamos ahora que $\lambda\partial f(x_0) \subset \partial(\lambda f)(x_0)$. Sea $l \in \partial f(x_0)$ y sea $\lambda > 0$, entonces

$$\lambda f(x) \geq \lambda f(x_0) + \lambda l(x - x0),$$

por tanto $\lambda l \in \partial(\lambda f)(x_0)$. Y se cumple $\partial(\lambda f)(x_0) = \lambda\partial f(x_0)$.

2

Sea $\|\cdot\|_\diamond$ una norma en $\mathbb{R}^n$ y sea

$$f(x) = \|x\|_\diamond.$$

Se define la norma dual del siguiente modo:

$$\|l\|_* = \max_{h\neq 0} \frac{|\langle l,h\rangle|}{\|h\|_\diamond} = \max_{\|h\|_\diamond=1} |\langle l,h\rangle|.$$

Pruebe que se cumplen las siguientes propiedades:

(a) $\partial f(0) = \{l \in \mathbb{R}^n : \|l\|_* \leq 1\}$. En particular, pruebe que

- (a.1) $\partial\|0\|_2 = \{l \in \mathbb{R}^n : \|l\|_2 \leq 1\}$,
- (a.2) $\partial\|0\|_1 = \{l \in \mathbb{R}^n : \|l\|_\infty \leq 1\}$,
- (a.3) $\partial\|0\|_\infty = \{l \in \mathbb{R}^n : \|l\|_1 \leq 1\}$,

(b) $\partial f(x) = \{l \in \mathbb{R}^n : \|l\|_* \leq 1,\ \langle l,x\rangle = \|x\|_\diamond\}$, para cada $x \in \mathbb{R}^n$.

(a)

Aplicando la definición de subgradiente tenemos

$$\begin{aligned}\partial\|0\|_\diamond &= \{l \in \mathbb{R}^n : \|x\|_\diamond \geq l(x),\ \forall x \in \mathbb{R}^n\}\\ &= \{l \in \mathbb{R}^n : \frac{|l(x)|}{\|x\|_\diamond} \leq 1,\ \forall x \in \mathbb{R}^n \setminus \{0\}\}\\ &= \{l \in \mathbb{R}^n : \|l\|_* \leq 1\}.\end{aligned}$$

Los siguientes subapartados son consecuencia del resultado anterior y que el dual de la norma ℓ^2 es ℓ^2, el dual de ℓ^1 es ℓ^∞ y el dual de ℓ^∞ es ℓ^1.

Norma dual de ℓ^2

Para ver que la norma dual de ℓ^2 es ella misma usamos la siguiente definición norma dual, para cualquier $y \in (\mathbb{R}^n)^* = \mathbb{R}^n$,

$$\|y\|_2^* = \sup_{\|x\|_2\leq 1} |\langle x,y\rangle|.$$

Usando la desigualdad de Cauchy-Schwarz tenemos,

$$|\langle x,y\rangle| \leq \|x\|_2\|y\|_2 \leq \|y\|_2,$$

, donde en la ultima desigualdad usamos $\|x\| \leq 1$. Finalmente si usamos $x = \frac{y}{\|y\|_2}$, que cumple $\|x\| = 1$, tenemos

$$|\langle x,y\rangle| = \|y\|_2,$$

alcanzando el supremo y por tanto

$$\|y\|_2^* = \|y\|_2.$$

Norma dual de ℓ^∞

Veamos que la norma dual de ℓ^∞ es ℓ^1, tenemos que para cualquier $y \in (\mathbb{R}^n)^* = \mathbb{R}^n$,

$$\|y\|_\infty^* = \sup_{\|x\|_\infty \leq 1} |\langle x, y\rangle| = \sup_{\max_{1\leq i\leq n} |x_i| \leq 1} \left| \sum_{i=1}^n x_i y_i \right|.$$

La restricción $\max_{1\leq i\leq n} |x_i| \leq 1$, implica que $x \in [-1,1]^n \subset \mathbb{R}^n$.

Para maximizar la suma elegimos $x \in [-1,1]^n$ tal que

$$x_i = \begin{cases} 1 \text{ si } y_i \geq 0, \\ -1 \text{ si } y_i < 0. \end{cases}$$

De este modo se tiene

$$|\langle x, y\rangle| = \sum_{i=1}^n |y_i|,$$

i por tanto

$$\|y\|_\infty^* = \sum_{i=1}^n |y_i| = \|y\|_1.$$

Norma dual de ℓ^1

Finalmente veamos que la norma dual de ℓ^1 es ℓ^∞, tenemos que para cualquier $y \in (\mathbb{R}^n)^* = \mathbb{R}^n$,

$$\|y\|_1^* = \sup_{\|x\|_1 \leq 1} |\langle x, y\rangle| = \sup_{\sum_{i=1}^n |x_i| \leq 1} \left| \sum_{i=1}^n x_i y_i \right|.$$

Para maximizar $|\langle x, y\rangle|$ con la restricción $\|x\|_1 \leq 1$, los componente $|x_i|$ deberían estar concentrados en maximizar los valores mas grandes $|y_i$. En particular el máximo se alcanza cuando $x_k = \text{signo}(y_k)$ para $|y_k|$ el componente mas grande de y y $x_i = 0$ para todo $i \neq k$. Esto es, sea $k = \text{argmax}_{1\leq i\leq n}|y_i|$, entonces

$$x_i = \begin{cases} \text{signo}(y_k) \text{ si } i = k, \\ 0 \text{ si } i \neq k. \end{cases}$$

En tal caso la norma dual es

$$\|y\|_1^* = \max_{1\leq i\leq n} |y_i| = \|y\|_\infty.$$

(b)

Veamos que

$$\partial f(x) = \{l \in \mathbb{R}^n : \|l\|_* \leq 1,\ \langle l, x\rangle = \|x\|_\diamond\},\ \forall x \in \mathbb{R}^n,$$

en el caso $x = 0$, ya hemos visto en el apartado anterior que se cumple, por tanto veamos que se cumple para $x \neq 0$.

Sea $l \in \mathbb{R}^n$ tal que $\|l\|_* \leq 1$ y $\langle l, x\rangle = \|x\|_\diamond$, tenemos que para cualquier $y \in \mathbb{R}^n$,

$$\begin{aligned}\|x\|_\diamond + \langle l, y - x\rangle &= \|x\|_\diamond + \langle l, y\rangle - \langle l, x\rangle \\ &= \|x\|_\diamond + \langle l, y\rangle - \|x\|_\diamond \\ &= \langle l, y\rangle \\ &\leq \|y\|_\diamond \cdot \frac{|\langle l, y\rangle|}{\|y\|_\diamond} \\ &\leq \|y\|_\diamond \|l\|_* \\ &\leq \|y\|_\diamond,\end{aligned}$$

por tanto $l \in \partial\|x\|_\diamond$, que implica $\{l \in \mathbb{R}^n : \|l\|_* \leq 1,\ \langle l, x\rangle = \|x\|_\diamond\} \subset \partial f(x)$, para todo $x \in \mathbb{R}^n$.

Sea $l \in \partial f(x)$, tenemos que

$$\|x\|_\diamond - \langle l, x\rangle = \|2x\|_\diamond - \|x\|_\diamond - \langle l, 2x - x\rangle \geq 0,$$

y

$$-\|x\|_\diamond + \langle l, x\rangle = \|0\|_\diamond - \|x\|_\diamond - \langle l, 0 - x\rangle \geq 0.$$

Las dos desigualdades llevan a $\|x\|_\diamond = \langle l, x\rangle$. Ademas para todo $y \in \mathbb{R}^n$, se cumple

$$\begin{aligned}\|y\|_\diamond &\geq \|x\|_\diamond + \langle l, y - x\rangle \\ &= \|x\|_\diamond + \langle l, y\rangle - \|x\|_\diamond \\ &= \langle l, y\rangle.\end{aligned}$$

Por lo tanto podemos concluir que

$$\|l\|_* = \sup_{y \in \mathbb{R}^n} \frac{|\langle l, y\rangle}{\|y\|_\diamond} \leq 1.$$

Por tanto se cumple $\partial f(x) \subset \{l \in \mathbb{R}^n : \|l\|_* \leq 1,\ \langle l, x\rangle = \|x\|_\diamond\}$, para todo $x \in \mathbb{R}^n$. Y con ambas inclusiones concluimos que $\partial f(x) = \{l \in \mathbb{R}^n : \|l\|_* \leq 1,\ \langle l, x\rangle = \|x\|_\diamond\}$.

3

Sean $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ dos funciones diferenciables y convexas. Se considera la función

$$f(x) = \max\{f1(x), f2(x)\}.$$

Sea $x_0 \in \mathbb{R}^n$ un punto tal que $f(x_0) = f_1(x_0) = f_2(x_0)$. Pruebe que l es un subgradiente de f en x_0 si y sólo si

$$l = \lambda \nabla f_1(x_0) + (1-\lambda)\nabla f_2(x_0), \text{ para } \lambda \in [0,1].$$

Supongamos que l es de la forma

$$l = \lambda \nabla f_1(x_0) + (1-\lambda)\nabla f_2(x_0),$$

con $\lambda \in [0,1]$. Entonces

$$\begin{aligned} f(x_0) + l(x-x_0) &= f(x_0) + \lambda \nabla f_1(x_0)(x-x_0) + (1-\lambda)\nabla f_2(x_0)(x-x_0) \\ &\leq f(x_0) + \lambda(f_1(x) - f_1(x_0)) + (1-\lambda)(f_2(x) - f_2(x_0)) \\ &= \lambda f_1(x) + (1-\lambda) f_2(x) \\ &\leq f(x), \end{aligned} \tag{1}$$

donde en la primera desigualdad hemos usado el Lema 3.3 del texto base para ver que

$$\nabla g(\hat{x})(x-\hat{x}) \leq g(x) - g(\hat{x}),$$

para toda función g convexa. Y la desigualdad (1) confirma que $l \in \partial f(x_0)$.

La idea para ver que todo subgradiente es de la forma

$$l = \lambda \nabla f_1(x_0) + (1-\lambda)\nabla f_2(x_0), \text{ para } \lambda \in [0,1],$$

es porque en el punto x_0 dado que las dos funciones coinciden sus dos vectores tangentes se cruzan en x_0, en cada dirección dominara uno u otro en función de si f_1 o f_2 es mayor que la otra en un pequeño entorno de x_0. Si el subgradiente no esta comprimido entre los dos vectores tangentes entonces en algún punto no se cumplirá la desigualdad $f(x) \geq f(x_0) + l(x-x_0)$. La Figura 1 ilustra esta idea en una dimension.

Vamos a formalizar la idea. Sea $l \in \partial f(x_0)$ tal que

$$l \neq l_\lambda = \lambda \nabla f_1(x_0) + (1-\lambda)\nabla f_2(x_0), \text{ para } \lambda \in [0,1].$$

Entonces existe $h \in \mathbb{R}^n$ tal que

$$l(h) > l_\lambda(h), \forall \lambda \in [0,1] \Rightarrow \begin{cases} l(h) > \nabla f_1(x_0)(h), \\ l(h) > \nabla f_2(x_0)(h). \end{cases}$$

Sin perdida de generalidad podemos suponer que para un $\alpha > 0$ suficientemente pequeño para que se cumpla

$$f(x_0 + \beta h) = f_1(x_0 + \beta h) \geq f_2(x_0 + \beta h), \ \forall 0 \leq \beta \leq \alpha.$$

Y por tanto en la dirección h la derivada direccional de f se corresponde con el gradiente de f_1,

$$f'(x_0)(h) = \nabla f_1(x_0)(h).$$

Entonces por el Lema 3.25 del texto base tenemos que

$$f'(x_0)(h) = \nabla f_1(x_0)(h) \geq l(h)!!$$

que contradice $l(h) > \nabla f_1(x_0)(h)$ y por tanto contradice $l \neq l_\lambda$. Resultando en que $\exists \lambda \in [0,1]$ tal que $l = l_\lambda$, como queríamos ver.

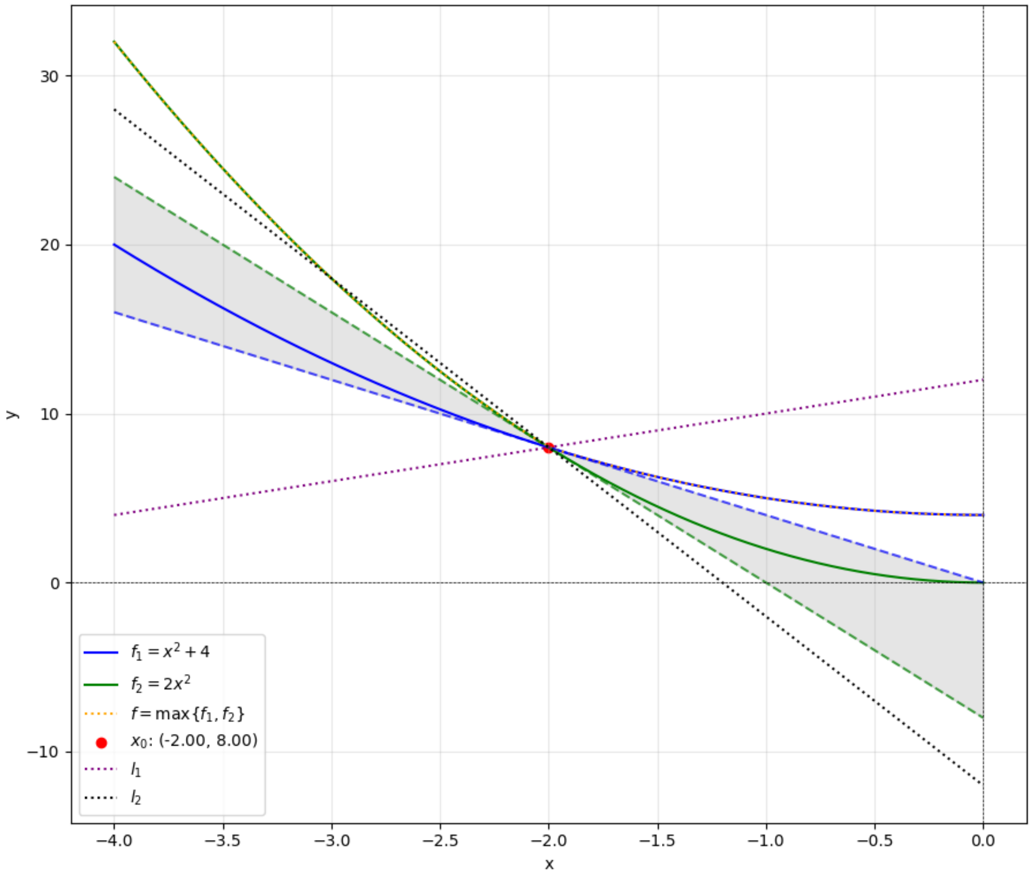

Figure 1: El gráfico muestra dos funciones convexas y su máximo, el area gris es el area comprimida entre los dos vectores tangentes, y l_1, l_2 son dos vectores que pasan por x_0 pero no están en el area gris, se ve como no cumplen la desigualdad.

4

Sea $(X, \|\cdot\|)$ un espacio real normado y $f : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$. La función $f^* : X^* \to \overline{\mathbb{R}}$ definida por

$$f^*(x^*) = \sup_{x \in X} \{\langle x, x^* \rangle - f(x)\},$$

se denomina la conjugada de Fenchel de f (X^* denota el conjunto de funcionales lineales y continuos $l : X \to \mathbb{R}$).

(a) Pruebe que

$$f(x) + f^*(x^*) \geq \langle x, x^* \rangle, \quad \forall x \in X, \quad x^* \in X^*.$$

La desigualdad anterior se conoce como *desigualdad de Young-Fenchel*.

(b) Sea $x_0 \in \mathrm{dom} f$. Demuestre que

$$x^* \in \partial f(x_0)$$

si y sólo si

$$f(x) + f^*(x^*) = \langle x_0, x^* \rangle, \tag{2}$$

y que lo anterior implica que $x_0 \in \partial f^*(x^*)$.

(c) Sea $x_0 \in \mathrm{dom} f$. Demuestre que

$$\partial f(x_0) \neq \emptyset \iff f(x_0) = \max_{x^* \in X^*} (\langle x_0, x^* \rangle - f^*(x^*)).$$

(a)

Para cualquier $x \in X$ y $x^* \in X^*$, tenemos

$$\begin{aligned} f(x) + f^*(x^*) &= f(x) + \sup_{\hat{x} \in X} \{\langle \hat{x}, x^* \rangle - f(\hat{x})\} \\ &\geq f(x) + \langle x, x^* \rangle - f(x) \\ &= \langle x, x^* \rangle. \end{aligned}$$

Como queríamos ver.

(b)

Supongamos que $x^* \in \partial f(x_0)$, entonces,

$$\begin{aligned} & f(x) \geq f(x_0) + \langle x - x_0, x^* \rangle, \ \forall x \in X, \\ & \quad \Rightarrow \langle x_0, x^* \rangle \geq f(x_0) + \{\langle x, x^* \rangle - f(x)\}, \ \forall x \in X, \\ & \quad \Rightarrow \langle x_0, x^* \rangle \geq f(x_0) + f^*(x^*). \end{aligned}$$

Usando el anterior resultado junto al resultado del apartado anterior, tenemos que

$$\langle x_0, x^* \rangle = f(x_0) + f^*(x^*).$$

Supongamos ahora que x^* cumple (2), entonces para todo $x \in X$, tenemos

$$\begin{aligned} f(x_0) + \langle x - x_0, x^* \rangle &= \langle x, x^* \rangle - \langle x_0, x^* \rangle + \langle x_0, x^* \rangle - f^*(x^*) \\ &= \langle x, x^* \rangle - \sup_{\hat{x} \in X} \{ \langle \hat{x}, x^* \rangle - f(\hat{x}) \} \\ &\leq \langle x, x^* \rangle - (\langle x, x^* \rangle - f(x)) \\ &= f(x), \end{aligned}$$

por tanto $x^* \in \partial f(x_0)$.

Finalmente veamos que si $x^* \in X^*$ y $x_0 \in X$ cumplen (2), entonces se tiene que para cualquier $y^* \in X^*$

$$\begin{aligned} f^*(x^*) + \langle x_0, y^* - x^* \rangle &= -f(x_0) + \langle x_0, x^* \rangle - \langle x_0, x^* \rangle + \langle x_0, y^* \rangle \\ &= \langle x_0, y^* \rangle - f(x_0) \\ &\leq \sup_{x \in X} \{ \langle x, y^* \rangle - f(x) \} \\ &= f^*(y^*), \end{aligned}$$

por tanto $x_0 \in \partial f^*(x^*)$.

(c)

Supongamos que $\partial f(x_0) \neq \emptyset$ y que $y^* \in \partial f(x_0)$, por la desigualdad de Young-Fenchel tenemos que

$$f(x_0) \geq \langle x_0, x^* \rangle - f^*(x^*), \ \forall x^* \in X^*,$$

lo que implica que

$$\begin{aligned} f(x_0) &\geq \max_{x^* \in X^*} (\langle x_0, x^* \rangle - f^*(x^*)) \\ &\geq \langle x_0, y^* \rangle - f^*(y^*) \\ &= f(x_0), \end{aligned} \tag{3}$$

donde en la ultima igualdad hemos usado el resultado del apartado anterior. Por tanto las desigualdades de (3) son igualdades y tenemos

$$f(x_0) = \max_{x^* \in X^*} (\langle x_0, x^* \rangle - f^*(x^*)).$$

Supongamos ahora que

$$f(x_0) = \max_{x^* \in X^*} (\langle x_0, x^* \rangle - f^*(x^*)).$$

Sea $y^* \in X^*$ tal que

$$\langle x_0, y^* \rangle - f^*(y^*) \geq \langle x_0, x^* \rangle - f^*(x^*),$$

es decir y^* es un máximo de la función $\langle x_0, x^* \rangle - f^*(x^*)$. Entonces tenemos que

$$f(x_0) + f^*(y^*) = \langle x_0, y^* \rangle,$$

por tanto usando el resultado del apartado anterior tenemos que $y^* \in \partial f(x_0)$, y por tanto $\partial f(x_0) \neq \emptyset$.