



Profesor: Fernando Roldán
 Ayudantes: Cristofer Mamani y Fernando Muñoz
 Departamento de Ingeniería Matemática
 Facultad de Ciencias Físicas y Matemáticas
 Universidad de Concepción

Optimización II (5225565)

Tarea 3

Fecha de entrega: Miércoles 19 de Noviembre de manera individual.

Para toda la tarea, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ y $g: \mathbb{R}^m \rightarrow \mathbb{R}$ son funciones convexas y denotamos $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^2: (x, y) \rightarrow (f(x), g(y))$.

P1) Pruebe que, para todo $x \in \mathbb{R}^n$ y todo $y \in \mathbb{R}^m$, se tiene $\partial F(x, y) = \partial f(x) \times \partial g(y)$.

Solución:

Sean $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^m$. Sea $(u, v) \in \partial F(x, y)$. se sigue que

$$\begin{aligned} & (\forall (z, w) \in \mathbb{R}^{n \times m}) \quad u^\top (x - z) + v^\top (y - x) + F(x, y) \leq F(w, z) \\ & \Leftrightarrow (\forall (z, w) \in \mathbb{R}^{n \times m}) \quad u^\top (x - z) + v^\top (y - x) + f(x) + g(y) \leq f(z) + g(w). \end{aligned} \tag{1}$$

Tomando $w = y$ se concluye $u \in \partial f(x)$ y tomando $z = x$ se concluye $v \in \partial g(y)$.

Por otro lado, dado $u \in \partial f(x)$ y $v \in \partial g(y)$, se tiene que

$$(\forall z \in \mathbb{R}^n) \quad u^\top (x - z) + f(x) \leq f(z).$$

y

$$(\forall w \in \mathbb{R}^n) \quad v^\top (y - w) + g(y) \leq g(w).$$

Sumando ambas desigualdades se tiene (1).

P2) Considere el operador proximal

$$\text{prox}_f: \mathbb{R}^n \rightarrow \mathbb{R}^n: x \rightarrow \arg \min_{y \in \mathbb{R}^n} \left(f(y) + \frac{1}{2} \|y - x\|^2 \right). \tag{2}$$

Demuestre que, para todo $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ y todo $\gamma > 0$:

- (i) $\text{prox}_{\gamma f}(x)$ existe y es único.
- (ii) $p = \text{prox}_{\gamma f}(x) \Leftrightarrow p = (\text{Id} + \gamma \partial f)^{-1}(x)$.
- (iii) $\text{prox}_{\gamma f}(x) = x - \gamma \text{prox}_{f^*/\gamma}(x/\gamma)$.
- (iv) $\|\text{prox}_{\gamma f}(x) - \text{prox}_{\gamma f}(y)\|^2 \leq \|x - y\|^2 - \|(x - \text{prox}_{\gamma f}(x)) - (y - \text{prox}_{\gamma f}(y))\|^2$.
- (v) $\text{prox}_{\gamma F}(x, z) = (\text{prox}_{\gamma f}(x), \text{prox}_{\gamma g}(z))$.

Solución:

- (i) Debido a que $x \mapsto \gamma f(x) + \frac{1}{2} \|y - x\|^2$ es fuertemente convexa, se tiene que $\arg \min_{y \in \mathbb{R}^n} (\gamma f(y) + \frac{1}{2} \|y - x\|^2)$ existe y es único.

(ii) Por la regla de Fermat se tiene que

$$\begin{aligned} p = \text{prox}_{\gamma f}(x) &\Leftrightarrow 0 \in \gamma \partial f(p) + p - x \\ &\Leftrightarrow x \in (\text{Id} + \gamma \partial f)(p) \\ &\Leftrightarrow p = (\text{Id} + \gamma \partial f)^{-1}(x), \end{aligned}$$

donde la ultima igualdad se sigue del hecho que p es único.

(iii) Se sabe que $\partial f^*(x) = (\partial f)^{-1}(x)$. Entonces

$$\begin{aligned} p = \text{prox}_{f^*/\gamma}(x/\gamma) &\Leftrightarrow x - \gamma p \in \partial f^*(p) \\ &\Leftrightarrow p \in \partial f(x - \gamma p) \\ &\Leftrightarrow x \in (\text{Id} + \gamma \partial f)(x - \gamma p) \\ &\Leftrightarrow (x - \gamma p) = \text{prox}_{\gamma f}(x), \end{aligned}$$

de donde se sigue el resultado.

(iv) Sea $p = \text{prox}_{\gamma f}(x)$ y $q = \text{prox}_{\gamma f}(y)$. Se sigue que $x - p \in \partial f(p)$, por lo tanto,

$$(\forall q \in \mathbb{R}^n) \quad (x - p)^\top (q - p) + f(p) \leq f(q).$$

Análogamente

$$(\forall p \in \mathbb{R}^n) \quad (y - q)^\top (p - q) + f(q) \leq f(p).$$

Sumando estas desigualdades se tiene que

$$\begin{aligned} (x - p - y + q)^\top (p - q) &\leq 0 \\ \Leftrightarrow \|x - p - (y - q)\|^2 + \|p - q\|^2 - \|x - y\|^2 &\leq 0. \end{aligned}$$

Demostrando lo pedido.

(v) Es directo combinando (ii) y **P1**.

P3) Sean $\rho > 0$ y $\gamma > 0$.

(i) Calcule $\text{prox}_{\gamma h_\rho}$ si $h_\rho: \mathbb{R} \rightarrow \mathbb{R}$ está dada por:

$$h_\rho(t) = \begin{cases} |t| - \frac{\rho}{2}, & \text{si } |t| > \rho; \\ \frac{t^2}{2\rho}, & \text{si } |t| \leq \rho. \end{cases} \quad (3)$$

(ii) Calcule $\text{prox}_{\gamma f}$ si $f: \mathbb{R}^n \rightarrow \mathbb{R}$ está dada por $f(x) = \sum_{i=1}^n h_\rho(x_i)$.

Solución:

(i) Note que h_ρ es diferenciable y

$$h'_\rho(t) = \begin{cases} 1, & \text{si } t > \rho \\ \frac{t}{\rho}, & \text{si } |t| \leq \rho \\ -1, & \text{si } t < -\rho. \end{cases} \quad (4)$$

Entonces,

$$\begin{aligned} p = \text{prox}_{\gamma h_\rho}(t) &\Leftrightarrow t - p = \gamma h'_\rho(p) \\ &\Leftrightarrow p = \begin{cases} t - \gamma, & \text{si } t > \rho + \gamma \\ \frac{\rho t}{\rho + \gamma}, & \text{si } |t| \leq \rho + \gamma \\ t + \gamma, & \text{si } t < -\rho - \gamma. \end{cases} \end{aligned}$$

(ii) Por **P2.(v)** extendido a la suma de n funciones, se sigue que

$$\text{prox}_{\gamma f}(x) = (\text{prox}_{\gamma h_\rho}(x_i))_{i=1}^n.$$

P4)

(i) Demuestre que $F^* = f^* + g^*$

(ii) Sea $p \in]1, +\infty[$ y defina $p^* = p/(p-1)$. Demuestre que $\left(\frac{|\cdot|^p}{p}\right)^* = \frac{|\cdot|^{p^*}}{p^*}$.

(iii) Demuestre que $\left(\frac{\|\cdot\|_p^p}{p}\right)^* = \frac{\|\cdot\|_{p^*}^{p^*}}{p^*}$.

(iv) Demuestre que $f = f^*$ si y solamente si $f^* = \|\cdot\|^2/2$.

Solución:

(i) Sea $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$. Se sigue que,

$$\begin{aligned} F^*(u, v) &= \sup_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} \langle (u, v) \mid (x, y) \rangle - F(x, y) \\ &= \sup_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} u^\top x + v^\top y - f(x) - f(y) \\ &= \sup_{x \in \mathbb{R}^n} u^\top x - f(x) + \sup_{y \in \mathbb{R}^m} v^\top y - f(y) \\ &= f^*(u) + g^*(v). \end{aligned}$$

(ii) Primero, definamos $f = \frac{|\cdot|^p}{p}$. Si $t > 0$, se sigue que $f'(t) = t^{p-1}$. Si $t < 0$, $f'(t) = -(-t)^{p-1} = -|t|^{p-1}$.

En general, si $t \neq 0$, $f'(t) = \text{sgn}(t)|t|^{p-1}$. Ahora, calculemos

$$f^*(u) = \sup_{t \in \mathbb{R}} g_u(t) := u \cdot t - f(t).$$

En el caso que $t \neq 0$, se tiene que $(g_u(t))' = u - \text{sgn}(t)|t|^{p-1}$. Este último término es 0 si y solo si

$$u = \text{sgn}(t)|t|^{p-1} \Leftrightarrow t = \text{sgn}(u)|u|^{\frac{1}{p-1}},$$

donde la última equivalencia se sigue notando que u y t tienen el mismo signo. Evaluando en $g_u(t)$, obtenemos

$$\begin{aligned} g_u(\text{sgn}(u)|u|^{\frac{1}{p-1}}) &= u \cdot \text{sgn}(u)|u|^{\frac{1}{p-1}} - \frac{|\text{sgn}(u)|u|^{\frac{1}{p-1}}|^p}{p} \\ &= |u|^{\frac{p}{p-1}} - \frac{|u|^{\frac{p}{p-1}}}{p} \\ &= \frac{|u|^{p^*}}{p^*} \geq 0. \end{aligned}$$

Debido a que $g_u(0) = 0$, se sigue el valor óptimo de $\sup_{t \in \mathbb{R}} g_u(t)$ se encuentra en que $t = \text{sgn}(u)|u|^{\frac{1}{p-1}}$ y por lo tanto $f^*(u) = \frac{|u|^{p^*}}{p^*}$.

(iii) Directo al combinar (i) (extendido a la suma de n funciones) con (ii).

- (iv) Si $f^* = \|\cdot\|^2/2$, por (iii) se sigue que $f = f^{**} = \|\cdot\|^2/2 = f^*$. Por otro lado, si $f = f^*$, por la desigualdad de Fenchel–Young

$$(\forall x \in \mathbb{R}^n) \quad f(x) + f^*(x) \geq \|x\|^2 \Rightarrow (\forall x \in \mathbb{R}^n) \quad f(x) \geq \|x\|^2/2.$$

Esto nos dice que $\|\cdot\|^2/2 = (\|\cdot\|^2/2)^* \geq f^* = f$, de donde se concluye que $f = \|\cdot\|^2/2$.

P5) Sean $u^i \in \mathbb{R}^n$ y $v^i \in \mathbb{R}$ para $i = 1, \dots, p$. Considere el problema

$$\begin{aligned} & \min_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}}} \quad \frac{\|x\|^2}{2} \\ & \text{s.a.} \quad v^i(x^\top u^i + y) \geq 1, \quad i = 1, \dots, p. \end{aligned}$$

- (i) Encuentre problema dual en el sentido de Lagrange.
(ii) Encuentre el problema dual en el sentido de Fenchel–Rockafellar.

Sean $u^i \in \mathbb{R}^n$ y $v^i \in \mathbb{R}$ para $i = 1, \dots, p$. Considere el problema

$$\begin{aligned} & \min_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}}} \quad \frac{\|x\|^2}{2} \\ & \text{s.a.} \quad v^i(x^\top u^i + y) \geq 1, \quad i = 1, \dots, p. \end{aligned}$$

Solución:

- (i) Notemos que la restricción es equivalente a

$$g_i(x, y) := 1 - v^i(u^{i\top} x + y) \leq 0.$$

Considerando multiplicadores $\mu_i \geq 0$, el Lagrangiano es

$$L(x, y, \mu) = \frac{1}{2}\|x\|^2 + \sum_{i=1}^p \mu_i(1 - v^i(u^{i\top} x + y)).$$

El problema dual es

$$\sup_{\mu_i \geq 0} \inf_{x, y} L(x, y, \mu).$$

Calculemos $\inf_{x, y} L(x, y, \mu)$. Las condiciones de primer orden son

$$\begin{aligned} \nabla_x L &= x - \sum_{i=1}^p \mu_i v^i u^i = 0 \quad \Rightarrow \quad x = \sum_{i=1}^p \mu_i v^i u^i, \\ \partial_y L &= - \sum_{i=1}^p \mu_i v^i = 0 \quad \Rightarrow \quad \sum_{i=1}^p \mu_i v^i = 0. \end{aligned}$$

Sustituyendo en L , se obtiene el problema dual:

$$\begin{aligned} & \sup_{\mu \in \mathbb{R}^p} \quad \sum_{i=1}^p \mu_i - \frac{1}{2} \left\| \sum_{i=1}^p \mu_i v^i u^i \right\|^2 \\ & \text{s.a.} \quad \mu_i \geq 0, \quad i = 1, \dots, p, \\ & \quad \sum_{i=1}^p \mu_i v^i = 0. \end{aligned}$$

(ii) Definiendo

$$\begin{aligned} f: \mathbb{R}^{n+1} &\rightarrow \mathbb{R}: (x, y) \mapsto \frac{1}{2} \|x\|^2 \\ A: \mathbb{R}^{n+1} &\rightarrow \mathbb{R}^p: (x, y) \mapsto (v^1(u^{1\top} x + y), \dots, v^p(u^{p\top} x + y)) \\ g: \mathbb{R}^p &\rightarrow \mathbb{R}: (z_i)_{i=1}^p \mapsto \sum_{i=1}^p \iota_{[1, +\infty[}(z_i), \end{aligned}$$

el problema primal se escribe como

$$\min_{(x, y) \in \mathbb{R}^{n+1}} f(x, y) + g(A(x, y)).$$

El dual de Fenchel–Rockafellar es

$$\min_{\mu \in \mathbb{R}^p} f^*(-A^\top \mu) + g^*(\mu).$$

En este caso, las conjugadas de Fenchel de f y g son:

$$\begin{aligned} f^*: \mathbb{R}^{n+1} &\rightarrow \mathbb{R}: (x, y) = \begin{cases} \frac{1}{2} \|x\|^2, & \text{si } y = 0, \\ +\infty, & \text{en otro caso.} \end{cases} \\ g^*: \mathbb{R}^p &\rightarrow \mathbb{R}: (z_i)_{i=1}^p \mapsto \begin{cases} \sum_{i=1}^p z_i, & \text{si } z_i \leq 0, \\ +\infty, & \text{en otro caso.} \end{cases} \end{aligned}$$

Ya que

$$A^\top: (\mu_i)_{i=1}^p \mapsto \left(\sum_{i=1}^p \mu_i v^i u^i, \sum_{i=1}^p \mu_i v^i \right),$$

tenemos que

$$f^*(-A^\top \mu) = \begin{cases} \frac{1}{2} \left\| \sum_{i=1}^p \mu_i v^i u^i \right\|^2, & \text{si } \sum_{i=1}^p \mu_i v^i = 0, \\ +\infty, & \text{en otro caso.} \end{cases}$$

Por tanto, el dual de Fenchel–Rockafellar es

$$\begin{aligned} \min_{\mu \in \mathbb{R}^p} \quad & \frac{1}{2} \left\| \sum_{i=1}^p \mu_i v^i u^i \right\|^2 + \sum_{i=1}^p \mu_i \\ \text{s.a.} \quad & \mu_i \leq 0, \quad i = 1, \dots, p, \\ & \sum_{i=1}^p \mu_i v^i = 0. \end{aligned}$$

P6) Dado $\alpha \in \mathbb{R}$, encuentre la solución del problema

$$\begin{aligned} \min_{(x, y, w, z) \in \mathbb{R}^4} \quad & x^2 + y^2 + w^2 + z^2 \\ \text{s.a.} \quad & x + y + w + z = 1 \\ & z \leq \alpha. \end{aligned}$$

Solución: Definamos

$$\begin{aligned} f: \mathbb{R}^4 &\rightarrow \mathbb{R}: (x, y, w, z) \mapsto x^2 + y^2 + w^2 + z^2 \\ h: \mathbb{R}^4 &\rightarrow \mathbb{R}: (x, y, w, z) \mapsto x + y + w + z - 1 \\ g: \mathbb{R}^4 &\rightarrow \mathbb{R}: (x, y, w, z) \mapsto z - \alpha. \end{aligned}$$

Se sigue que f es fuertemente convexa (note que $f = \|\cdot\|_{\mathbb{R}^4}^2$), g es convexa, y h es afín. Adicionalmente, el problema se puede escribir como

$$\begin{aligned} & \min_{(x,y,w,z) \in \mathbb{R}^4} f(x, y, w, z) \\ \text{s.a. } & h(x, y, w, z) = 0 \\ & g(x, y, w, z) \leq 0. \end{aligned}$$

Notemos que, por ejemplo, el vector $(x_0, y_0, w_0, z_0) = (0, 0, 2 - \alpha, \alpha - 1)$ cumple la condición de Slater. Entonces, por las condiciones de KKT, un vector (x, y, w, z) es solución del problema, si y solo si, existe $\lambda \geq 0$ y $\mu \in \mathbb{R}$ tales que

$$\begin{aligned} 0 &= \nabla f(x, y, z, w) + \lambda \nabla g(x, y, z, w) + \mu \nabla h(x, y, z, w) \\ 0 &\geq g(x, y, z, w) \\ 0 &= \lambda \cdot g(x, y, z, w) \\ 0 &= h(x, y, z, w), \end{aligned}$$

lo que en este caso se traduce a

$$\begin{aligned} (0, 0, 0, 0) &= (2x, 2y, 2w, 2z) + \lambda(0, 0, 0, 1) + \mu(1, 1, 1, 1) \\ 0 &\geq z - \alpha \\ 0 &= \lambda \cdot (z - \alpha) \\ 0 &\leq \lambda \\ 0 &= x + y + w + z - 1. \end{aligned}$$

Se sigue que $x = y = w = -\mu/2$ y $z = -(\lambda + \mu)/2$. Ahora,

$$\begin{aligned} x + y + w + z - 1 = 0 &\Leftrightarrow 2\mu + \frac{\lambda}{2} = -1 \Leftrightarrow \mu = -\frac{\lambda + 2}{4} \\ &\Rightarrow x = y = w = \frac{\lambda + 2}{8} \quad y \quad z = \frac{2 - 3\lambda}{8}. \end{aligned}$$

Por lo tanto, necesitamos que

$$\frac{2 - 3\lambda}{8} \leq \alpha.$$

- En el caso que $\lambda = 0$, se tiene $\mu = -1/2$ y $(x, y, z, w) = \frac{1}{4}(1, 1, 1, 1)$, siempre que $\alpha \geq 1/4$.
- Si $\lambda > 0$, se tiene que $\alpha = z = \frac{2-3\lambda}{8} \Leftrightarrow \lambda = \frac{2-8\alpha}{3}$ y entonces $\lambda > 0 \Leftrightarrow \alpha < \frac{1}{4}$. Se sigue que $\mu = \frac{2\alpha-2}{3}$ y $(x, y, w) = -\frac{\mu}{2} = \frac{1-\alpha}{3}$.

En resumen se presenta la siguiente tabla.

α	(x, y, w, z)	λ	μ	$f(x, y, w, z)$
$\left[\frac{1}{4}, +\infty\right]$	$\frac{1}{4}(1, 1, 1, 1)$	0	$-\frac{1}{2}$	$\frac{1}{4}$
$\left]-\infty, \frac{1}{4}\right]$	$\left(\frac{1-\alpha}{3}, \frac{1-\alpha}{3}, \frac{1-\alpha}{3}, \alpha\right)$	$\frac{2-8\alpha}{3}$	$\frac{2\alpha-2}{3}$	$\frac{1}{3}(1 - 2\alpha^2 + 4\alpha^2)$