PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
Escuela de ingeniería
ICS3153 Optimización Avanzada
Profesor: Cristóbal Guzmán (crguzmanp@uc.cl)
Alumno: Alberto Valdés (anvaldes@uc.cl)

Tarea 4

Pregunta 1

Solución a):

Lo primero que haremos será definir la función $h : E \times E \to \bar{\mathbb{R}}$

$$R(x,y) = f(y) + \frac{1}{2\lambda}\|x-y\|_{\mathbf{E}}^2$$

Partiremos demostrando que la función $g(x,y) : E \times E \to \bar{\mathbb{R}}$ definida por:

$$g(x,y) = \|x-y\|_{\mathbf{E}} = \|x-y\|_{\mathbf{E}}$$

es convexa sobre $E \times E$.

Sean $z_1 = (x_1,y_1), z_2 = (x_2,y_2) \in E \times E$ arbitrarios con $x_1,x_2,y_1,y_2 \in E$ y $t \in [0,1]$ también arbitrario. Notemos que:

$$g(t \cdot z_1 + [1-t] \cdot z_2) = g(t \cdot (x_1,y_1) + [1-t] \cdot (x_2,y_2)) = g(t \cdot x_1 + [1-t] \cdot x_2, t \cdot y_1 + [1-t] \cdot y_2)$$

$$= \|(t \cdot x_1 + [1-t] \cdot x_2) - (t \cdot y_1 + [1-t] \cdot y_2)\|_{\mathbf{E}} = \|t \cdot (x_1-y_1) + [1-t] \cdot (x_2-y_2)\|_{\mathbf{E}}$$

$$\overset{\text{desigualdad triangular}}{\leq} \|t \cdot (x_1-y_1)\|_{\mathbf{E}} + \|[1-t] \cdot (x_2-y_2)\|_{\mathbf{E}} = |t| \cdot \|x_1-y_1\|_{\mathbf{E}} + |1-t| \cdot \|x_2-y_2\|_{\mathbf{E}}$$

$$\overset{t\geq 0,(1-t)\geq 0}{=} t \cdot \|x_1-y_1\|_{\mathbf{E}} + (1-t) \cdot \|x_2-y_2\|_{\mathbf{E}} = t \cdot g(x_1,y_1) + (1-t) \cdot g(x_2,y_2) = t \cdot g(z_1) + (1-t) \cdot g(z_2)$$

$\Rightarrow$

$$g(t \cdot z_1 + [1-t] \cdot z_2) \leq t \cdot g(z_1) + (1-t) \cdot g(z_2)$$

Y como $z_1, z_2 \in E \times E$ son arbitrarios y $t \in [0,1]$ también lo es, entonces:

$$g(t \cdot z_1 + [1-t] \cdot z_2) \leq t \cdot g(z_1) + (1-t) \cdot g(z_2) \quad (\forall t \in [0,1], \forall z_1, z_2 \in E \times E)$$

$$\Rightarrow \boxed{g \text{ es convexa } (1)}$$

Ahora sea la función $w(u) = \dfrac{u^2}{2\lambda}$. Notemos que $w'(u) = \dfrac{2 \cdot u}{2 \cdot \lambda} = \dfrac{u}{\lambda}$, por lo tanto $\forall u \geq 0$ se cumple que $w'(u) \geq 0$ por ende $w(u)$ es creciente $\forall u \geq 0$. Además también notemos que $w''(u) = \dfrac{1}{\lambda} > 0$ por ende:

$$\boxed{w \text{ es convexa y además } \forall t \geq 0 \; w \text{ es creciente } (2)}$$

Ahora notemos que si definimos la función $l(x,y) = (wog)(x,y) = w(g(x,y))$, entonces:

$$l(x,y) = w(g(x,y)) = \frac{[g(x,y)]^2}{2\lambda} = \frac{\|x-y\|^2_{\mathbf{E}}}{2\lambda} \Rightarrow \boxed{l(x,y) = \frac{\|x-y\|^2_{\mathbf{E}}}{2\lambda} \; (3)}$$

De esta forma, sean $z_1 = (x_1, y_1), z_2 = (x_2, y_2) \in E \times E$ arbitrarios con $x_1, x_2, y_1, y_2 \in E$ y $t \in [0,1]$ también arbitrario. Notemos que:

$$\boxed{l(t \cdot z_1 + [1-t] \cdot z_2) = w[g(t \cdot z_1 + [1-t] \cdot z_2)] \; (4)}$$

Ahora como g es convexa sobre $E \times E$, entonces:

$$g(t \cdot z_1 + [1-t] \cdot z_2) \leq \; t \cdot g(z_1) + [1-t] \cdot g(z_2)$$

Como sabemos que g es siempre positiva pues $g(x,y) = \|x-y\|^2_{\mathbf{E}} \geq 0$ y como $w(u)$ es creciente para todo $u \geq 0$, entonces:

$$\boxed{w(g(t \cdot z_1 + [1-t] \cdot z_2)) \leq \; w(t \cdot g(z_1) + [1-t] \cdot g(z_2)) \; (5)}$$

Usando (4) y (5) llegamos a que:

$$l(t \cdot z_1 + [1-t] \cdot z_2) = w[g(t \cdot z_1 + [1-t] \cdot z_2)] \leq w(g(t \cdot z_1 + [1-t] \cdot z_2)) \leq \; w(t \cdot g(z_1) + [1-t] \cdot g(z_2))$$

$$\overset{w \text{ es convexa}}{\leq} t \cdot w(g(z_1)) + [1-t] \cdot w(g(z_2)) \overset{definicion\ de\ l}{=} t \cdot l(z_1) + [1-t] \cdot l(z_2)$$

De esta forma:

$$l(t \cdot z_1 + [1-t] \cdot z_2) \leq t \cdot l(z_1) + [1-t] \cdot l(z_2)$$

Y como se tomo $z_1, z_2 \in E \times E$ arbitrarios y $t \in [0,1]$ también arbitrario, entonces:

$$l(t \cdot z_1 + [1-t] \cdot z_2) \leq t \cdot l(z_1) + [1-t] \cdot l(z_2)(\forall t \in [0,1])(\forall z_1, z_2 \in E \times E)$$

Lo cual implica que:

$$\boxed{l(x,y) = w(g(x,y)) = \frac{\|x-y\|_{\mathbb{E}}^2}{2\lambda} \text{ es convexa sobre } E \times E \quad (6)}$$

Ahora definiremos la función $R : E \times E :\to \bar{\mathbb{R}}$ tal que:

$$R(x,y) = f(y) + \frac{1}{2\lambda}\|x-y\|_{\mathbb{E}}^2 = f(y) + l(x,y)$$

Como por enunciado tenemos que f es convexa y como por (6) tenemos que $l(x,y)$ es convexa sobre $E \times E$ entonces sean $z_1 = (x_1,y_1), z_2 = (x_2,y_2) \in E \times E$ arbitrarios con $x_1,x_2,y_1,y_2 \in E$ y sea $t \in [0,1]$ también arbitrario, entonces:

$$R(t\cdot z_1+[1-t]\cdot z_2) = R(t\cdot(x_1,y_1)+[1-t]\cdot(x_2,y_2)) = R(t\cdot x_1+[1-t]\cdot x_2, t\cdot y_1+[1-t]\cdot y_2) \overset{definicion\ de\ R}{=}$$

$$f(t\cdot y_1+[1-t]\cdot y_2) + l(t\cdot x_1+[1-t]\cdot x_2, t\cdot y_1+[1-t]\cdot y_2) \overset{convexidad\ de\ f\ sobre\ E}{\leq}$$

$$t\cdot f(y_1)+[1-t]\cdot f(y_2)+l(t\cdot x_1+[1-t]\cdot x_2, t\cdot y_1+[1-t]\cdot y_2) =$$

$$t\cdot f(y_1)+[1-t]\cdot f(y_2)+l(t\cdot z_1+[1-t]\cdot z_2) \overset{convexidad\ de\ l\ sobre\ E\times E}{\leq}$$

$$t\cdot f(y_1)+[1-t]\cdot f(y_2)+t\cdot l(z_1)+[1-t]\cdot l(z_2) = t\cdot[f(y_1)+l(z_1)]+[1-t]\cdot[f(y_2)+l(z_2)]$$

$$\overset{z_1=(x_1,y_1),z_2=(x_2,y_2)}{=} t\cdot[f(y_1)+l(x_1,y_1)]+[1-t]\cdot[f(y_2)+l(x_2,y_2)]$$

$$\overset{Definicion\ de\ R}{=} t\cdot R(x_1,y_1)+[1-t]\cdot R(x_2,y_2) \overset{z_1=(x_1,y_1),z_2=(x_2,y_2)}{=} t\cdot R(z_1)+[1-t]\cdot R(z_2)$$

$\Rightarrow$

$$R(t\cdot z_1+[1-t]\cdot z_2) \leq t\cdot R(z_1)+[1-t]\cdot R(z_2)$$

Como $z_1, z_2 \in E \times E$ son arbitrarios y $t \in [0,1]$ también es arbitrario, entonces:

$$\boxed{R(t\cdot z_1+[1-t]\cdot z_2) \leq t\cdot R(z_1)+[1-t]\cdot R(z_2)\ (\forall t \in [0,1])(\forall z_1,z_2 \in E\times E) \quad (7)}$$

Por lo tanto R es convexa sobre $E \times E$.

Notemos que:

$$f_\lambda(x) = \inf_{y\in E} R(x,y)$$

Ahora, sean $x_1, x_2 \in E$ arbitrarios y $t \in [0,1]$ arbitrario, entonces:

$$f_\lambda(t \cdot x_1 + [1-t] \cdot x_2) = \inf_{y \in E} R(t \cdot x_1 + [1-t] \cdot x_2, y)$$

$$= \inf_{y_1, y_2 \in E} R(t \cdot x_1 + [1-t] \cdot x_2, t \cdot y_1 + [1-t] \cdot y_2) \overset{Convexidad\ de\ R}{\leq} \inf_{y_1, y_2 \in E} [t \cdot R(x_1, y_1) + [1-t] \cdot R(x_2, y_2)]$$

$$\overset{Problemas\ separables}{=} \inf_{y_1 \in E}(t \cdot R(x_1, y_1)) + \inf_{y_2 \in E}([1-t] \cdot R(x_2, y_2))$$

$$= t \cdot \inf_{y_1 \in E} R(x_1, y_1) + [1-t] \cdot \inf_{y_2 \in E} R(x_2, y_2) = t \cdot f_\lambda(x_1) + [1-t] \cdot f_\lambda(x_2)$$

$\Rightarrow$

$$f_\lambda(t \cdot x_1 + [1-t] \cdot x_2) \leq t \cdot f_\lambda(x_1) + [1-t] \cdot f_\lambda(x_2)$$

Como se demostró para $x_1, x_2 \in E$ arbitrarios y para $t \in [0,1]$ arbitrario, entonces:

$$f_\lambda(t \cdot x_1 + [1-t] \cdot x_2) \leq t \cdot f_\lambda(x_1) + [1-t] \cdot f_\lambda(x_2) \quad (\forall t \in [0,1], \forall x_1, x_2 \in E)$$

$\Rightarrow$

$$\boxed{f_\lambda \text{ es convexa} \quad (8)}$$

Ahora notemos que:

$$f_\lambda(x) = \min_{y \in E}\left[f(y) + \frac{1}{2\lambda}\|x-y\|_{\mathbb{E}}^2\right] \leq f(u) + \frac{1}{2\lambda}\|x-u\|_{\mathbb{E}}^2 \ \forall u \in \mathbb{R}^n$$

Y en especifico esto se cumple para $u = x$, de este modo:

$$f_\lambda(x) \leq f(x) + \frac{1}{2\lambda}\|x-x\|_{\mathbb{E}}^2 = f(x)$$

$\Rightarrow$

$$\boxed{f_\lambda(x) \leq f(x) \quad (9)}$$

Demostrando así lo pedido.

Notemos que como $f_\lambda(x) = \inf_{y \in E}\left[f(y) + \dfrac{1}{2\lambda}\|x-y\|_{\mathbf{E}}^2\right]$, entonces:

$$\forall \epsilon > 0, \exists y_\epsilon \in E : f(y_\epsilon) + \frac{1}{2\lambda}\|x-y_\epsilon\|_{\mathbf{E}}^2 - f_\lambda(x) \leq \epsilon$$

En particular, podemos tomar $\epsilon = \lambda > 0$. De este modo:

$$\exists y_\lambda \in E : f(y_\lambda) + \frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2 - f_\lambda(x) \leq \lambda$$

$\Rightarrow$

$$\boxed{f(y_\lambda) + \frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2 \leq \lambda + f_\lambda(x) \overset{por\ (9)}{\leq} \lambda + f(x) \quad (10)}$$

Ahora como $f \in \Gamma_0(E)$, entonces existe $g \in \partial f(x)$, y por definición de $\partial f(x)$ tenemos que:

$$f(x) + \langle g, v - x\rangle \leq f(v) \ \forall v \in E$$

Y en particular tomamos $v = y_\lambda$, de este modo:

$$\boxed{f(x) + \langle g, y_\lambda - x\rangle \leq f(y_\lambda) \quad (11)}$$

Uniendo (10) y (11) llegamos a que:

$$f(x) + \langle g, y_\lambda - x\rangle + \frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2 \leq f(y_\lambda) + \frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2 \leq \lambda + f_\lambda(x) \leq \lambda + f(x)$$

$\Rightarrow$

$$\boxed{f(x) + \langle g, y_\lambda - x\rangle + \frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2 \leq \lambda + f_\lambda(x) \leq \lambda + f(x) \quad (12)}$$

Ahora, esto implica que:

$$f(x) + \langle g, y_\lambda - x\rangle + \frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2 \leq \lambda + f(x)$$

$\Rightarrow$

$$\langle g, y_\lambda - x\rangle + \frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2 \leq \lambda$$

$\Rightarrow$

$$2\lambda \cdot \langle g, y_\lambda - x\rangle + \|x - y_\lambda\|_{\mathbf{E}}^2 \leq 2\lambda \cdot \lambda$$

$\Rightarrow$

$$2\langle \lambda g, y_\lambda - x\rangle + \|x - y_\lambda\|_{\mathbf{E}}^2 \leq 2\lambda^2$$

$\Rightarrow$

$$\boxed{2\langle \lambda g, y_\lambda - x\rangle + \|x - y_\lambda\|_{\mathbf{E}}^2 \leq 2\lambda^2 \quad (13)}$$

$\Rightarrow$

$$\|\lambda g\|_{\mathbf{E}}^2 + 2\langle \lambda g, y_\lambda - x\rangle + \|x - y_\lambda\|_{\mathbf{E}}^2 \leq 2\lambda^2 + \|\lambda g\|_{\mathbf{E}}^2$$

$\Rightarrow$

$$\|\lambda g + y_\lambda - x\|_{\mathbf{E}}^2 \leq 2\lambda^2 + \lambda^2 \cdot \|g\|_{\mathbf{E}}^2$$

$\Rightarrow$

$$\|y_\lambda - [x - \lambda g]\|_{\mathbf{E}}^2 \leq 2\lambda^2 + \lambda^2 \cdot \|g\|_{\mathbf{E}}^2$$

$\Rightarrow$

$$\|y_\lambda - [x - \lambda g]\|_{\mathbf{E}} \leq \sqrt{2\lambda^2 + \lambda^2 \cdot \|g\|_{\mathbf{E}}^2} = \lambda\sqrt{2 + \|g\|_{\mathbf{E}}^2}$$

$\Rightarrow$

$$y_\lambda \in B_{\|\cdot\|_{\mathbf{E}}}\left[x - \lambda g, \lambda\sqrt{2 + \|g\|_{\mathbf{E}}^2}\right]$$

De este modo, es claro que si $\lambda \to 0^+$, entonces:

$$\lim_{\lambda \to 0^+} y_\lambda \in B_{\|\cdot\|_{\mathbf{E}}}[x, 0] = \{x\}$$

Así tenemos que:

$$\boxed{\lim_{\lambda \to 0^+} y_\lambda = x \quad (14)}$$

Usando esto en (12) y tomando limite de cuando $\lambda \to 0^+$ tenemos que:

$$\lim_{\lambda \to 0^+}\left[f(x) + \langle g, y_\lambda - x\rangle + \frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2\right] \leq \lim_{\lambda \to 0^+}[\lambda + f_\lambda(x)] \leq \lim_{\lambda \to 0^+}[\lambda + f(x)]$$

$\Rightarrow$

$$f(x) + \lim_{\lambda \to 0^+}\langle g, y_\lambda - x\rangle + \lim_{\lambda \to 0^+}\left(\frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2\right) \leq \lim_{\lambda \to 0^+} f_\lambda(x) \leq f(x)$$

Como la función $g(u) = \langle g, u - x\rangle$ es afín, entonces es continua, de este modo:

$$f(x) + \langle g, \left(\lim_{\lambda \to 0^+} y_\lambda\right) - x\rangle + \lim_{\lambda \to 0^+}\left(\frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2\right) \leq \lim_{\lambda \to 0^+} f_\lambda(x) \leq f(x)$$

Y usando (14) tenemos que:

$$\boxed{f(x) + \lim_{\lambda \to 0^+}\left(\frac{1}{2\lambda}\|x - y_\lambda\|_{\mathbf{E}}^2\right) \leq \lim_{\lambda \to 0^+} f_\lambda(x) \leq f(x) \quad (15)}$$

Ahora notemos que por (13) tenemos que:

$$2\lambda\langle g, y_\lambda - x\rangle + \|x - y_\lambda\|_{\mathbf{E}}^2 \leq 2\lambda^2$$

$\Rightarrow$

$$\langle g, y_\lambda - x\rangle + \frac{1}{2\lambda}\cdot\|x - y_\lambda\|_{\mathbf{E}}^2 \leq \lambda$$

Ahora tiramos limite cuando $\lambda \to 0^+$, entonces:

$$\lim_{\lambda\to 0^+}\langle g, y_\lambda - x\rangle + \lim_{\lambda\to 0^+}\left(\frac{1}{2\lambda}\cdot\|x - y_\lambda\|_{\mathbf{E}}^2\right) \leq \lim_{\lambda\to 0^+}\lambda$$

$\Rightarrow$

$$\lim_{\lambda\to 0^+}\langle g, y_\lambda - x\rangle + \lim_{\lambda\to 0^+}\left(\frac{1}{2\lambda}\cdot\|x - y_\lambda\|_{\mathbf{E}}^2\right) \leq 0$$

Como sabemos que $g(u) = \langle g, u - x\rangle$ es afin, por ende continua, entonces:

$$\langle g, \left(\lim_{\lambda\to 0^+} y_\lambda\right) - x\rangle + \lim_{\lambda\to 0^+}\left(\frac{1}{2\lambda}\cdot\|x - y_\lambda\|_{\mathbf{E}}^2\right) \leq 0$$

Usando (14) tenemos que:

$$\boxed{\lim_{\lambda\to 0^+}\left(\frac{1}{2\lambda}\cdot\|x - y_\lambda\|_{\mathbf{E}}^2\right) \leq 0 \quad (16)}$$

Ahora notemos que como las normas son siempre positivas y como $\lambda > 0$, entonces:

$$\frac{1}{2\lambda}\cdot\|x - y_\lambda\|_{\mathbf{E}}^2 \geq 0 \ (\forall\lambda > 0)$$

$\Rightarrow$

$$\boxed{\lim_{\lambda\to 0^+}\left(\frac{1}{2\lambda}\cdot\|x - y_\lambda\|_{\mathbf{E}}^2\right) \geq 0 \quad (17)}$$

Usando (16) y (17) llegamos a que:

$$\boxed{\lim_{\lambda\to 0^+}\left(\frac{1}{2\lambda}\cdot\|x - y_\lambda\|_{\mathbf{E}}^2\right) = 0 \quad (18)}$$

Reemplazando esto en (15) tenemos que:

$$\boxed{f(x) \leq \lim_{\lambda\to 0^+} f_\lambda(x) \leq f(x) \quad (19)}$$

Lo cual implica que:

$$\boxed{\lim_{\lambda\to 0^+} f_\lambda(x) = f(x) \quad (20)}$$

Demostrando así lo pedido.

Así usando (8), (9) y (20) tenemos que demostramos todo lo pedido en este inciso [Pregunta 1. a)]

Solución b):

Notemos que:

$$(f\Box g)^*(y) = \sup_{x\in E}\left[\langle y,x\rangle - (f\Box g)(x)\right] = \sup_{x\in E}\left[\langle y,x\rangle - \inf_{z\in E}[f(z)+g(x-z)]\right] =$$

$$\sup_{x\in E}\left[\langle y,x\rangle + \sup_{z\in E}[-f(z)-g(x-z)]\right] \overset{Como\ \langle y,x\rangle\ no\ depende\ de\ z}{=} \sup_{x\in E}\sup_{z\in E}[\langle y,x\rangle - f(z) - g(x-z)]$$

$$\overset{Podemos\ intercambiar\ supremos}{=} \sup_{z\in E}\ \sup_{x\in E}[\langle y,x\rangle - f(z) - g(x-z)] = \sup_{z\in E}\ \sup_{x\in E}[\langle y, x-z+z\rangle - f(z) - g(x-z)]$$

$$= \sup_{z\in E}\ \sup_{x\in E}[\langle y,x-z\rangle + \langle y,z\rangle - f(z) - g(x-z)] = \sup_{z\in E}\ \left(\sup_{x\in E}[\langle y,x-z\rangle + \langle y,z\rangle - f(z) - g(x-z)]\right)$$

$$\overset{Como\ \langle y,z\rangle - f(z)\ no\ depende\ de\ x}{=} \sup_{z\in E}\ \left(\sup_{x\in E}[\langle y,x-z\rangle - g(x-z)] + \langle y,z\rangle - f(z)\right)$$

$$\overset{Tomando\ a=x-z\ y\ que\ [\sup_{x\in E}=\sup_{a\in E}]}{=} \sup_{z\in E}\ \left(\sup_{a\in E}[\langle y,a\rangle - g(a)] + \langle y,z\rangle - f(z)\right) = \sup_{z\in E}\,[g^*(y)+\langle y,z\rangle - f(z)]$$

$$\overset{Como\ g^*(y)\ no\ depende\ de\ z}{=} \sup_{z\in E}[\langle y,z\rangle - f(z)] + g^*(y) = f^*(y) + g^*(y)$$

$\Rightarrow$

$$\boxed{(f\Box g)^*(y) = f^*(y) + g^*(y) \quad (21)}$$

Ahora notemos que como $f_\lambda(x) = \left(f\Box\frac{1}{2\lambda}\|\cdot\|^2_{\mathbf{E}}\right)(x)$ y usando (21) llegamos a que:

$$(f_\lambda)^*(y) = \left(f\Box\frac{1}{2\lambda}\|\cdot\|^2_{\mathbf{E}}\right)^*(y) \overset{Por\ (21)}{=} f^*(y) + \left(\frac{1}{2\lambda}\|\cdot\|^2_{\mathbf{E}}\right)^*(y)$$

$\Rightarrow$

$$\boxed{(f_\lambda)^*(y) = f^*(y) + \left(\frac{1}{2\lambda}\|\cdot\|^2_{\mathbf{E}}\right)^*(y) \quad (22)}$$

Ahora consideremos $g(x) = \frac{1}{2\lambda}\|x\|^2_{\mathbf{E}}$, entonces:

$$g^*(y) = \sup_{x\in E}[\langle y,x\rangle - g(x)] = \sup_{x\in E}\left[\langle y,x\rangle - \frac{1}{2\lambda}\|x\|^2_{\mathbf{E}}\right] = \sup_{r>0}\left(\sup_{u\in E:\|u\|_{\mathbf{E}}=1}\left[\langle y,ru\rangle - \frac{1}{2\lambda}\|ru\|^2_{\mathbf{E}}\right]\right)$$

$$= \sup_{r>0} \left(\sup_{u \in E: \|u\|_{\mathbf{E}}=1} \left[r \cdot \langle y, u \rangle - r^2 \cdot \frac{1}{2\lambda} \|u\|_{\mathbf{E}}^2 \right] \right) = \sup_{r>0} \left(\sup_{u \in E: \|u\|_{\mathbf{E}}=1} \left[r \cdot \langle y, u \rangle - r^2 \cdot \frac{1}{2\lambda} \right] \right)$$

$$= \sup_{r>0} \left(r \cdot \left[\sup_{u \in E: \|u\|_{\mathbf{E}}=1} \langle y, u \rangle \right] - r^2 \cdot \frac{1}{2\lambda} \right) \overset{Por\ definicion\ de\ norma\ dual}{=} \sup_{r>0} \left(r \cdot \|y\|_{\mathbf{E},*} - r^2 \cdot \frac{1}{2\lambda} \right)$$

$$\overset{norma\ dual\ de\ \|\|_{\mathbf{E}}\ es\ \|\|_{\mathbf{E}}}{=} \sup_{r>0} \left(r \cdot \|y\|_{\mathbf{E}} - r^2 \cdot \frac{1}{2\lambda} \right)$$

$\Rightarrow$

$$\boxed{g^*(y) = \sup_{r>0} \left(r \cdot \|y\|_{\mathbf{E}} - r^2 \cdot \frac{1}{2\lambda} \right) \quad (23)}$$

Sea $k(r) = r \cdot \|y\|_{\mathbf{E}} - r^2 \cdot \frac{1}{2\lambda}$, notemos que:

$$k'(r) = \|y\|_{\mathbf{E}} - \frac{r}{\lambda}$$

Y además que:

$$k''(r) = -\frac{1}{\lambda} < 0$$

Por lo tanto, como siempre $k''(r) < 0$ entonces k es concava, por ende el punto r' que llega al supremo es el punto tal que $k'(r') = 0$. Ahora pasaremos a encontrar ese r':

$$k'(r') = 0 \ \Rightarrow \ \|y\|_{\mathbf{E}} - \frac{r'}{\lambda} = 0 \ \Rightarrow \ \frac{r'}{\lambda} = \|y\|_{\mathbf{E}} \ \Rightarrow \ r' = \lambda \cdot \|y\|_{\mathbf{E}}$$

De este modo, combinando esto con (23) llegamos a que:

$$g^*(y) = \sup_{r>0} \left(r \cdot \|y\|_{\mathbf{E}} - r^2 \cdot \frac{1}{2\lambda} \right) = \sup_{r>0} (k(r)) = k(r') = k(\lambda \cdot \|y\|_{\mathbf{E}}) = \lambda \cdot \|y\|_{\mathbf{E}} \cdot \|y\|_{\mathbf{E}} - \lambda^2 \cdot \|y\|_{\mathbf{E}}^2 \cdot \frac{1}{2\lambda}$$

$$= \lambda \cdot \|y\|_{\mathbf{E}} \cdot \|y\|_{\mathbf{E}} - \frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2} = \lambda \cdot \|y\|_{\mathbf{E}}^2 - \frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2} = \frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2}$$

Así reemplazando esto en (22) llegamos a que:

$$(f_\lambda)^*(y) = f^*(y) + \left(\frac{1}{2\lambda} \|\cdot\|_{\mathbf{E}}^2 \right)^* (y) = f^*(y) + \frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2}$$

$\Rightarrow$

$$\boxed{(f_\lambda)^*(y) = f^*(y) + \frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2} \quad (24)}$$

Sea $q(y) = \frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2}$, notemos que:

$$\nabla q(y) = \lambda y \ \Rightarrow \ \nabla^2 q(y) = \lambda \cdot I$$

Como sabemos, para este caso en particular decimos que su constante de σ fuerte convexidad esta dada por el menor valor propio de $\nabla^2 g(y)$, pero como todos sus valores propios son iguales a λ entonces decimos que $\sigma = \lambda$, de este modo $\frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2}$ es $\lambda-$ fuertemente convexa.

Ahora usando esto y que por definición $f^*(y)$ es convexa, y recordando (24), entonces tomamos $y_1, y_2 \in E$ arbitrarios y $t \in [0,1]$ arbitrario también, entonces:

$$(f_\lambda)^*(t \cdot y_1 + [1-t] \cdot y_2) = f^*(t \cdot y_1 + [1-t] \cdot y_2) + \frac{\lambda \cdot \|t \cdot y_1 + [1-t] \cdot y_2\|_{\mathbf{E}}^2}{2}$$

$$\overset{Como\ q(y)=\frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2}}{=} f^*(t \cdot y_1 + [1-t] \cdot y_2) + q(t \cdot y_1 + [1-t] \cdot y_2) \overset{convexidad\ de\ f^*}{\leq} t \cdot f^*(y_1) + [1-t] \cdot f^*(y_2) + q(t \cdot y_1 + [1-t] \cdot y_2)$$

$$\overset{\lambda-fuerte\ convexidad\ de\ q}{\leq} t \cdot f^*(y_1) + [1-t] \cdot f^*(y_2) + t \cdot q(y_1) + [1-t] \cdot q(y_2) - \frac{\lambda t(1-t)}{2} \cdot \|y_1 - y_2\|_{\mathbf{E}}^2$$

$$\overset{reagrupando\ terminos}{=} t \cdot [f^*(y_1) + q(y_1)] + [1-t] \cdot [f^*(y_2) + q(y_2)] - \frac{\lambda t(1-t)}{2} \cdot \|y_1 - y_2\|_{\mathbf{E}}^2$$

$$\overset{como\ (f_\lambda)^*(y)=f^*(y)+q(y)}{=} t \cdot (f_\lambda)^*(y_1) + [1-t] \cdot (f_\lambda)^*(y_2) - \frac{\lambda t(1-t)}{2} \cdot \|y_1 - y_2\|_{\mathbf{E}}^2$$

$\Rightarrow$

$$(f_\lambda)^*(t \cdot y_1 + [1-t] \cdot y_2) \leq t \cdot (f_\lambda)^*(y_1) + [1-t] \cdot (f_\lambda)^*(y_2) - \frac{\lambda \cdot t \cdot (1-t)}{2} \cdot \|y_1 - y_2\|_{\mathbf{E}}^2$$

Como esto se cumple para $y_1, y_2 \in E$ arbitrarios y como $t \in [0,1]$ también es arbitrario, entonces:

$$(f_\lambda)^*(t \cdot y_1 + [1-t] \cdot y_2) \leq t \cdot (f_\lambda)^*(y_1) + [1-t] \cdot (f_\lambda)^*(y_2) - \frac{\lambda \cdot t \cdot (1-t)}{2} \cdot \|y_1 - y_2\|_{\mathbf{E}}^2 \quad (\forall t \in [0,1], \forall y_1, y_2 \in E)$$

$\Rightarrow$

$$\boxed{(f_\lambda)^* \text{ es } \lambda - \text{fuertemente convexa respecto a } \|\|_{\mathbf{E}} \ \ (25)}$$

Ahora, recordando que por (24) tenemos que:

$$(f_\lambda)^*(y) = f^*(y) + \frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2}$$

Y como $f \in \Gamma_0(E)$, además de que:

$$\left(\frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2} > -\infty \ \ \forall y \in E \ \text{ ademas de que } \ \forall y \in E \ \ \frac{\lambda \cdot \|y\|_{\mathbf{E}}^2}{2} < +\infty \ \Rightarrow \ dom\left((f_\lambda)^*\right) \neq \emptyset \right)$$

entonces $(f_\lambda)^*$ es propia $[(f_\lambda)^*(y) > -\infty \ \forall y \in E$ y $dom((f_\lambda)^*) \neq \emptyset]$ y como por definición de conjugada de fenchel se tiene que $(f_\lambda)^*$ es semi continua inferior, entonces por teorema dual entre $\sigma-$fuerte convexidad y $\mu-$suavidad, tenemos que $[(f_\lambda)^*]^* = (f_\lambda)^{**}$ es $\frac{1}{\lambda}$ suave respecto a $\|\cdot\|_{\mathbf{E},*} = \|\cdot\|_{\mathbf{E}}$, entonces:

$$\boxed{(f_\lambda)^{**} \text{ es } \frac{1}{\lambda}\text{-suave respecto a } \|\cdot\|_{\mathbf{E}} \quad (26)}$$

que es lo mismo a decir que:

$$\boxed{(f_\lambda)^{**} \in \mathcal{F}^1_{\mathbf{E},\|\cdot\|_{\mathbf{E}}}\left(\frac{1}{\lambda}\right) \quad (27)}$$

Ahora recordando que:

$$f_\lambda(x) = \inf_{y\in E}\left[f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right]$$

Ahora definamos la función $U^x(y) = f(y) + \dfrac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}$ con x fijo. Notemos que si definimos $H^x(y) = \dfrac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}$, entonces es claro que:

$$\nabla H^x(y) = \frac{1}{\lambda}(y-x) \;\Rightarrow\; \nabla^2 H^x(y) = \frac{1}{\lambda}\cdot I$$

Recordemos que para este caso en particular se tiene que H^x es $\sigma-$fuertemente convexa de parametro igual al menor valor propio de $\nabla^2 H^x(y)$ y como $\nabla^2 H^x(y) = \frac{1}{\lambda}\cdot I$ entonces todos sus valores propios son iguales a $\frac{1}{\lambda}$ y por ende H^x es $\frac{1}{\lambda}$-fuertemente convexa, de este modo usando esto y que f es convexa, entonces sean $y_1, y_2 \in E$ arbitrarios y $t \in [0,1]$ también arbitrario, entonces:

$$U^x(t\cdot y_1 + [1-t]\cdot y_2) = f(t\cdot y_1 + [1-t]\cdot y_2) + \frac{1}{2\lambda}\|x - [t\cdot y_1 + [1-t]\cdot y_2]\|^2_{\mathbf{E}}$$

$$\overset{Como\ H^x(y)=\frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}}{=} f(t\cdot y_1 + [1-t]\cdot y_2) + H^x(t\cdot y_1 + [1-t]\cdot y_2)$$

$$\overset{convexidad\ de\ f}{\leq} t\cdot f(y_1) + [1-t]\cdot f(y_2) + H^x(t\cdot y_1 + [1-t]\cdot y_2)$$

$$\overset{\frac{1}{\lambda}-fuerte\ convexidad\ de\ H^x}{\leq} t\cdot f(y_1) + [1-t]\cdot f(y_2) + t\cdot H^x(y_1) + [1-t]\cdot H^x(y_2) - \frac{1}{2\lambda}\cdot t\cdot(1-t)\cdot\|y_1-y_2\|^2_{\mathbf{E}}$$

$$\overset{reagrupando\ terminos}{=} t\cdot[f(y_1) + H^x(y_1)] + [1-t]\cdot[f(y_2) + H^x(y_2)] - \frac{1}{2\lambda}\cdot t\cdot(1-t)\cdot\|y_1-y_2\|^2_{\mathbf{E}}$$

$$\overset{como\ U^x(y)=f(y)+H^x(y)}{=} t\cdot U^x(y_1) + [1-t]\cdot U^x(y_2) - \frac{1}{2\lambda}\cdot t\cdot(1-t)\cdot\|y_1-y_2\|^2_{\mathbf{E}}$$

$\Rightarrow$

$$U^x(t\cdot y_1 + [1-t]\cdot y_2) \leq t\cdot U^x(y_1) + [1-t]\cdot U^x(y_2) - \frac{1}{2\lambda}\cdot t\cdot(1-t)\cdot\|y_1-y_2\|^2_{\mathbf{E}}$$

Como se tomo $y_1, y_2 \in E$ arbitrarios y $t \in [0,1]$ también arbitrario, entonces:

$$U^x(t\cdot y_1+[1-t]\cdot y_2) \leq t\cdot U^x(y_1)+[1-t]\cdot U^x(y_2)-\frac{1}{2\lambda}\cdot t\cdot(1-t)\cdot\|y_1-y_2\|_{\mathbf{E}}^2 \ \ (\forall t\in[0,1], \forall y_1, y_2 \in E)$$

entonces U^x es $\frac{1}{\lambda}$-fuertemente convexa, por lo tanto, siempre existe un único minimizador y como:

$$f_\lambda(x) = \inf_{y\in E} U^x(y)$$

entonces $\forall x \in E, \exists y_x \in E$ único tal que:

$$f_\lambda(x) = U(y_x) \overset{como\ U^x(y)=f(y)+H^x(y)}{=} f(y_x)+H^x(y_x) = f(y_x)+\frac{1}{2\lambda}\|x-y_x\|_{\mathbf{E}}^2$$

$\Rightarrow$

$$f_\lambda(x) = f(y_x)+\frac{1}{2\lambda}\|x-y_x\|_{\mathbf{E}}^2$$

Y como f es propia, entonces $f(y_x) > -\infty$ y es claro que $\frac{1}{2\lambda}\|x-y_x\|_{\mathbf{E}}^2 > -\infty$, de este modo:

$$f_\lambda(x) > -\infty$$

y como se demostro esto para un x arbitrario, entonces:

$$\boxed{f_\lambda(x) > -\infty \quad \forall x \in E \quad (28)}$$

Y como tenemos que por $a)$ que $f_\lambda(x) \leq f(x)$ y como $f \in \Gamma_0(E)$ entonces f es propia, entonces $dom(f) \neq \emptyset$, entonces existe $v \in dom(f)$ y así tenemos que:

$$f_\lambda(v) \leq f(v) \overset{como\ v\in dom(f)}{<} +\infty \ \Rightarrow \ f_\lambda(v) < +\infty$$

De este modo, $v \in dom(f_\lambda)$ así tenemos que:

$$\boxed{dom(f_\lambda) \neq \emptyset \quad (29)}$$

Ahora usando (28) y (29) entonces:

$$\boxed{f_\lambda \ \text{ es propia} \quad (30)}$$

Ahora usando lo que demostramos más arriba de que $\forall x \in E, \exists$ un único y_x tal que:

$$f_\lambda(x) = f(y_x)+\frac{1}{2\lambda}\|y_x-x\|_{\mathbf{E}}^2$$

Entonces consideremos $(x_n)_{n\in\mathbb{N}} \subseteq E$ arbitraria tal que $x_n \to \bar{x}$ para un $\bar{x} \in E$ arbitrario. Es claro que $y_{x_n} \to y_{\bar{x}}$, de este modo:

$$f_\lambda(\bar{x}) = f(y_{\bar{x}})+\frac{1}{2\lambda}\|y_{\bar{x}}-\bar{x}\|_{\mathbf{E}}^2 \overset{como\ f\ es\ sci}{\leq} \liminf_{n\to+\infty} f(y_{x_n})+\frac{1}{2\lambda}\|y_{\bar{x}}-\bar{x}\|_{\mathbf{E}}^2$$

$$= \liminf_{n\to+\infty} f(y_{x_n})+\frac{1}{2\lambda}\|\lim_{n\to+\infty}[y_{x_n}-x_n]\|_{\mathbf{E}}^2 \overset{continuidad\ de\ \frac{1}{2\lambda}\|x-y\|_{\mathbf{E}}^2}{=} \liminf_{n\to+\infty} f(y_{x_n})+\lim_{n\to+\infty}\frac{1}{2\lambda}\|y_{x_n}-x_n\|_{\mathbf{E}}^2$$

$$\overset{como\ \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\ es\ continua\ \ \liminf=\lim}{=} \liminf_{n\to+\infty} f(y_{x_n}) + \liminf_{n\to+\infty} \frac{1}{2\lambda}\|y_{x_n} - x_n\|^2_{\mathbf{E}}$$

$$\overset{como\ sumas\ de\ \liminf\ es\ menor\ que \liminf\ de\ la\ suma}{\leq} \liminf_{n\to+\infty}\left[f(y_{x_n}) + \frac{1}{2\lambda}\|y_{x_n} - x_n\|^2_{\mathbf{E}}\right] = \liminf_{n\to+\infty} f_\lambda(x_n)$$

De esta forma:

$$f_\lambda(\bar{x}) \leq \liminf_{n\to+\infty} f_\lambda(x_n)$$

Y como esto se demostro para una sucesión arbitraria $(x_n)_{n\in\mathbb{N}} \subseteq E : x_n \to \bar{x}$ entonces se demostró para toda sucesión. Y ahora como se demostró para un $\bar{x}$ arbitrario, entonces se demostró para todo $\bar{x} \in E$, de este modo, se demostró que:

$$\boxed{f_\lambda \text{ es semi continua inferior } \ (31)}$$

Ahora tomando (30), (31) y el resultado de a) que dice que f_λ es convexa, entonces:

$$\boxed{f_\lambda \in \Gamma_0(E) \quad (32)}$$

De este modo $f_\lambda = (f_\lambda)^{**}$ y usando (27) llegamos a que:

$$\boxed{f_\lambda \in \mathcal{F}^1_{\mathbf{E},\|\cdot\|_{\mathbf{E}}}\left(\frac{1}{\lambda}\right) \quad (33)}$$

Así de esta forma, se demostró todo lo pedido en el inciso b) de la pregunta 1.

Solución c):

Notemos que como C es convexo y cerrado, entonces $\chi_C(x) \in \Gamma_0(E)$. De esta modo:

$$f_\lambda(x) = \inf_{y\in E}\left[\chi_C(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right] = \inf_{y\in C}\left[\frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right] = \frac{1}{2\lambda}\inf_{y\in C}\left[\|x-y\|^2_{\mathbf{E}}\right]$$

$$\overset{definicion\ de\ proyeccion}{=} \frac{1}{2\lambda}\|x - \Pi_C(x)\|^2_{\mathbf{E}}$$

Por lo tanto:

$$\boxed{f_\lambda(x) = \frac{1}{2\lambda}\|x - \Pi_C(x)\|^2_{\mathbf{E}} \text{ para } f(x) = \chi_C(x)}$$

que es justamente lo que nos pedían calcular.

Pregunta 2

Solución a):

Primero que todo definiremos la función $U^x(y)$ para un $x \in E$ fijo. Esta función estará dada por:

$$U^x(y) = f(y) + \frac{\|x-y\|_{\mathbf{E}}^2}{2} = f(y) + H^x(y)$$

con:

$$H^x(y) = \frac{\|x-y\|_{\mathbf{E}}^2}{2}$$

Notemos que:

$$\nabla H^x(y) = (y-x) \ \Rightarrow \ \nabla^2 H^x(y) = I$$

Decimos que $H^x(y)$ es σ-fuertemente convexa de parámetro igual al menor valor propio del hessiano, y como el hessiano esta dado por la matriz identidad entonces todos los valores propios de la matriz hessiana valen 1 y por ende $H^x(y)$ es 1-fuertemente convexa.

De este modo, sean $y_1, y_2 \in E$ arbitrarios y $t \in [0,1]$ también arbitrarios, entonces:

$$U^x(t \cdot y_1 + [1-t] \cdot y_2) = f(t \cdot y_1 + [1-t] \cdot y_2) + H^x(t \cdot y_1 + [1-t] \cdot y_2)$$

$$\overset{Como\ f \in \Gamma_0(E) \Rightarrow f\ es\ convexa}{\leq} t \cdot f(y_1) + [1-t] \cdot f(y_2) + H^x(t \cdot y_1 + [1-t] \cdot y_2)$$

$$\overset{H^x(y)\ es\ 1-fuertemente\ convexa}{\leq} t \cdot f(y_1) + [1-t] \cdot f(y_2) + t \cdot H^x(y_1) + [1-t] \cdot H^x(y_2) - \frac{1}{2} \cdot t \cdot (1-t) \cdot \|y_1 - y_2\|_{\mathbf{E}}^2$$

$$\overset{Reagrupando\ terminos}{=} t \cdot (f(y_1) + H^x(y_1)) + [1-t] \cdot (f(y_2) + H^x(y_2)) - \frac{1}{2} \cdot t \cdot (1-t) \cdot \|y_1 - y_2\|_{\mathbf{E}}^2$$

$$\overset{Como\ U^x(y) = f(y) + H^x(y)}{=} t \cdot U^x(y_1) + [1-t] \cdot U^x(y_2) - \frac{1}{2} \cdot t \cdot (1-t) \cdot \|y_1 - y_2\|_{\mathbf{E}}^2$$

$\Rightarrow$

$$U^x(t \cdot y_1 + [1-t] \cdot y_2) \leq t \cdot U^x(y_1) + [1-t] \cdot U^x(y_2) - \frac{1}{2} \cdot t \cdot (1-t) \cdot \|y_1 - y_2\|_{\mathbf{E}}^2$$

Como se demostró para $t \in [0,1]$ arbitrario y para $y_1, y_2 \in E$ arbitrarios también, entonces:

$$U^x(t \cdot y_1 + [1-t] \cdot y_2) \leq t \cdot U^x(y_1) + [1-t] \cdot U^x(y_2) - \frac{1}{2} \cdot t \cdot (1-t) \cdot \|y_1 - y_2\|_{\mathbf{E}}^2 \ \ (\forall t \in [0,1], \forall y_1, y_2 \in E)$$

Y esta es la definición de una función 1-fuertemente convexa, por lo tanto:

$$\boxed{U^x(y) \ \text{ es 1-fuertemente convexa} \quad (1)}$$

De este modo, tenemos que:

$$\boxed{\text{Existe un único minimizador de } U^x(y) \quad (2)}$$

Y como tenemos que:

$$Prox_f(x) = \min_{y \in E} U^x(y)$$

entonces:

$$\boxed{Prox_f \text{ es una función, es decir que para todo } x \text{ el argmin es un singleton.} \quad (3)}$$

Ahora consideremos a $z = Prox_f(x)$. Notemos que esto significa que $0 \in \partial U^x(z)$.

Como $U^x(y) = f(y) + \frac{\|x-y\|_{\mathbf{E}}^2}{2}$, entonces:

$$\partial U^x(y) = \partial f(y) + \{y - x\}$$

Como tenemos que $0 \in \partial U^x(z)$, entonces:

$$0 \in \partial f(z) + \{z - x\}$$

$\Rightarrow$

$$x \in \partial f(z) + \{z\} = (\partial + I)(z)$$

$\Rightarrow$

$$(\partial + I)^{-1}(x) \in \{z\} \overset{habiamos\ dicho\ que\ z = Prox_f(x)}{=} \{Prox_f(x)\}$$

Como acabamos de demostrar que $Prox_f(x)$ es una función, entonces $\{Prox_f(x)\}$ es un singleton.

Por ende, como $(\partial + I)^{-1}(x) \in \{Prox_f(x)\}$, entonces:

$$\boxed{(\partial + I)^{-1}(x) = Prox_f(x) \quad (4)}$$

que es justamente lo que nos estaban pidiendo demostrar.

Solución b):

i)

$$[x^* \in argmin\{f(x) : x \in E] \overset{Teorema\ de\ fermat}{\Leftrightarrow} [0 \in \partial f(x^*)] \Leftrightarrow [x^* \in \partial f(x^*) + \{x^*\} = (\partial + I)(x^*)]$$

$$\Leftrightarrow [(\partial{+}I)^{-1}(x^*) \in \{x^*\}] \overset{Por\ 2a)}{\Leftrightarrow} [Prox_f(x^*) \in \{x^*\}] \Leftrightarrow [Prox_f(x^*) = x^*] \Leftrightarrow [x^* \text{ es punto fijo de } Prox_f]$$

De este modo, se demostró que:

$$\boxed{[x^* \in argmin\{f(x) : x \in E] \Leftrightarrow [x^* \text{ es punto fijo de } Prox_f] \quad (5)}$$

que es justamente lo que nos estaban pidiendo demostrar.

---

ii)

Demostraremos esto por casos.

<u>Caso 1) $Prox_f(x) = Prox_f(y)$:</u> En este caso la desigualdad se demuestra trivialmente pues:

$$\|Prox_f(x) - Prox_f(y)\|_{\mathbf{E}} = \|0\|_{\mathbf{E}} = 0 \overset{Por\ definicion\ de\ norma}{\leq} \|x - y\|_{\mathbf{E}}$$

$\Rightarrow$

$$\boxed{\|Prox_f(x) - Prox_f(y)\|_{\mathbf{E}} = \|0\|_{\mathbf{E}} \leq \|x - y\|_{\mathbf{E}}}$$

que es justamente lo que estábamos buscando demostrar.

<u>Caso 2) $Prox_f(x) \neq Prox_f(y)$:</u>

Sea $z = Prox_f(x)$, entonces por el resultado de la $P2a)$ tenemos que $z = (I + \partial f)^{-1}(x)$.

De esta forma:

$$(I + \partial f)^{-1}(x) \in \{z\} \;\Rightarrow\; x \in (I + \partial f)(z) = \partial f(z) + \{z\} \;\Rightarrow\; \boxed{x - z \in \partial f(z)}$$

Notemos que por definición de subgradiente tenemos que:

$$\boxed{f(z) + \langle x - z, u - z\rangle \leq f(u) \ \forall u \in E \quad (6)}$$

Por otro lado:

Sea $w = Prox_f(y)$, entonces por el resultado de la $P2a)$ tenemos que $w = (I + \partial f)^{-1}(y)$.

De esta forma:

$$(I+\partial f)^{-1}(y) \in \{w\} \;\Rightarrow\; y \in (I+\partial f)(w) = \partial f(w) + \{w\} \;\Rightarrow\; \boxed{y - w \in \partial f(w)}$$

Notemos que por definición de subgradiente tenemos que:

$$\boxed{f(w) + \langle y-w, v-w\rangle \leq f(v)\ \forall v \in E \quad (7)}$$

Usando (6) en el caso particular en el que $u = w$ obtenemos que:

$$\boxed{f(z) + \langle x-z, w-z\rangle \leq f(w) \quad (8)}$$

Y usando (7) en el caso particular en el que $v = z$ obtenemos que:

$$\boxed{f(w) + \langle y-w, z-w\rangle \leq f(z) \quad (9)}$$

Sumando (8) y (9) llegamos a que:

$$f(z) + \langle x-z, w-z\rangle + f(w) + \langle y-w, z-w\rangle \leq f(w) + f(z)$$

$\Rightarrow$

$$\langle x-z, w-z\rangle + \langle y-w, z-w\rangle \leq 0$$

$\Rightarrow$

$$\langle x-z, w-z\rangle - \langle y-w, w-z\rangle \leq 0$$

$\Rightarrow$

$$\langle x-z-[y-w], w-z\rangle \leq 0$$

$\Rightarrow$

$$\langle x-z-y+w, w-z\rangle \leq 0$$

$\Rightarrow$

$$\langle x-y, w-z\rangle + \langle w-z, w-z\rangle \leq 0$$

$\Rightarrow$

$$\langle x-y, w-z\rangle \leq -\langle w-z, w-z\rangle = -\|w-z\|_{\mathbf{E}}^2$$

$\Rightarrow$

$$\langle x-y, w-z\rangle \leq -\|w-z\|_{\mathbf{E}}^2$$

$\Rightarrow$

$$\|w-z\|_{\mathbf{E}}^2 \leq \langle x-y, w-z\rangle \overset{Definicion\ de\ valor\ absoluto}{\leq} |\langle x-y, w-z\rangle| \overset{Cauchy-Schwarz}{\leq} \|x-y\|_{\mathbf{E}}\|w-z\|_{\mathbf{E}}$$

$\Rightarrow$

$$\|w-z\|_{\mathbf{E}}^2 \leq \|x-y\|_{\mathbf{E}}\|w-z\|_{\mathbf{E}}$$

Reemplazando con que $z = Prox_f(x)$ y con que $w = Prox_f(y)$, entonces:

$$\|Prox_f(y) - Prox_f(x)\|_{\mathbf{E}}^2 \leq \|x-y\|_{\mathbf{E}}\|Prox_f(y) - Prox_f(x)\|_{\mathbf{E}}$$

Como estamos en el caso en que $Prox_f(y) \neq Prox_f(x)$, entonces $\|Prox_f(y) - Prox_f(x)\| > 0$ y por ende podemos dividir por tal norma la desigualdad anterior, obteniendo que:

$$\boxed{\|Prox_f(y) - Prox_f(x)\| \leq \|x-y\|_{\mathbf{E}}}$$

Así como se demostró esta desigualdad para el caso $Prox_f(y) = Prox_f(x)$ y para el caso en que $Prox_f(y) \neq Prox_f(x)$, entonces se demostró para todo caso. De esta forma, tenemos que:

$$\boxed{\|Prox_f(y) - Prox_f(x)\| \leq \|x-y\|_{\mathbf{E}} \quad (10)}$$

que es justamente lo que nos estaban pidiendo demostrar.

---

iii)

Sea $z = Prox_f(x)$, entonces por $P2a)$ tenemos que $(I+\partial f)^{-1}(x) = z$ lo cual implica que:

$$(I+\partial f)^{-1}(x) \in \{z\} \;\Rightarrow\; x \in (I+\partial f)(z) = \{z\} + \partial f(z) \;\Rightarrow\; \boxed{x - z \in \partial f(z)}$$

Por desigualdad de fenchel esto implica que:

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

Como $f \in \Gamma_0(E)$, lo que implica que $f^{**} = f$, entonces tenemos que:

$$f^{**}(z) + f^*(x-z) = \langle x-z, x\rangle$$

$\Rightarrow$

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

Esto implica que $z \in \partial f^*(x-z)$. Así tenemos que:

$$z + (x-z) \in \partial f^*(x-z) + \{x-z\} = (\partial f^* + I)(x-z)$$

$\Rightarrow$

$$x \in (\partial f^* + I)(x-z) \;\Rightarrow\; (I+\partial f^*)^{-1}(x) \in \{x-z\} \;\Rightarrow\; (I+\partial f^*)^{-1}(x) = x - z$$

Luego, por definición de $Prox_{f^*}$ dada en $P2a)$ tenemos que:

$$Prox_{f^*}(x) = x - z \;\Rightarrow\; z + Prox_{f^*}(x) = x$$

Y como dijimos en un inicio que $z = Prox_f(x)$ entonces tenemos que:

$$\boxed{Prox_f(x) + Prox_{f^*}(x) = x \quad (11)}$$

Que es justamente lo que estábamos buscando demostrar.

Solución c):

Notemos que si $f(x) = \sum_{j=1}^{d} f_j(x_j)$ entonces:

$$f_\lambda(x) = \inf_{y\in\mathbb{R}^d}\left[f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right] = \inf_{y\in\mathbb{R}^d}\left[\sum_{j=1}^{d} f_j(x_j) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right]$$

$$\overset{expansion\ de\ norma\ 2\ al\ cuadrado}{=} \inf_{y\in\mathbb{R}^d}\left[\sum_{j=1}^{d} f_j(x_j) + \frac{1}{2\lambda}\sum_{j=1}^{d}(x_j-y_j)^2\right]$$

$$= \inf_{y\in\mathbb{R}^d}\left[\sum_{j=1}^{d} f_j(x_j) + \frac{1}{2\lambda}\sum_{j=1}^{d}|x_j-y_j|^2\right] = \inf_{y\in\mathbb{R}^d}\left[\sum_{j=1}^{d} f_j(x_j) + \sum_{j=1}^{d}\frac{1}{2\lambda}|x_j-y_j|^2\right]$$

$$= \inf_{y\in\mathbb{R}^d}\left[\sum_{j=1}^{d}\left(f_j(x_j) + \frac{1}{2\lambda}|x_j-y_j|^2\right)\right] = \inf_{y_1,\ldots,y_d\in\mathbb{R}}\left[\sum_{j=1}^{d}\left(f_j(x_j) + \frac{1}{2\lambda}|x_j-y_j|^2\right)\right]$$

$$\overset{problema\ separable}{=} \sum_{j=1}^{d}\left[\inf_{y_j\in\mathbb{R}}\left(f_j(x_j) + \frac{1}{2\lambda}|x_j-y_j|^2\right)\right]$$

$\Rightarrow$

$$\boxed{\inf_{y\in\mathbb{R}^d}\left[f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right] = \sum_{j=1}^{d}\left[\inf_{y_j\in\mathbb{R}}\left(f_j(x_j) + \frac{1}{2\lambda}|x_j-y_j|^2\right)\right] \quad (12)}$$

$\Rightarrow$

$$f_\lambda(x) = \inf_{y\in\mathbb{R}^d}\left[f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right] = \sum_{j=1}^{d}\left[\inf_{y_j\in\mathbb{R}}\left(f_j(x_j) + \frac{1}{2\lambda}|x_j-y_j|^2\right)\right] \overset{por\ definicion\ de\ (f_j)_\lambda(x_j)}{=} \sum_{j=1}^{d}(f_j)_\lambda(x_j)$$

$\Rightarrow$

$$\boxed{f_\lambda(x) = \sum_{j=1}^{d}(f_j)_\lambda(x_j) \quad (13)}$$

que es justamente uno de los puntos que nos estaban pidiendo demostrar.

Ahora notemos que:

$$Prox_{\lambda f}(x) = argmin_{y\in\mathbb{R}^d}\left[\lambda f(y) + \frac{1}{2}\|x-y\|^2_{\mathbf{E}}\right] = argmin_{y\in\mathbb{R}^d}\left[\lambda\left(f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right)\right]$$

$$\overset{como\ \lambda>0\ y\ por\ propiedades\ del\ argmin}{=} argmin_{y\in\mathbb{R}^d}\left[f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right]$$

$\Rightarrow$

$$\boxed{Prox_{\lambda f}(x) = argmin_{y\in\mathbb{R}^d}\left[f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right] \quad (14)}$$

Recordando que $f(x) = \sum_{j=1}^d f_j(x_j)$, de este modo:

$$f(y)+\frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}} = \sum_{j=1}^d f_j(x_j)+\frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}} \overset{expansion\ de\ norma\ 2\ al\ cuadrado}{=} \sum_{j=1}^d f_j(x_j)+\frac{1}{2\lambda}\sum_{j=1}^d(x_j-y_j)^2$$

$$= \sum_{j=1}^d f_j(x_j) + \frac{1}{2\lambda}\sum_{j=1}^d|x_j-y_j|^2 = \sum_{j=1}^d f_j(x_j) + \sum_{j=1}^d \frac{1}{2\lambda}|x_j-y_j|^2 = \sum_{j=1}^d\left(f_j(x_j) + \frac{1}{2\lambda}|x_j-y_j|^2\right)$$

$\Rightarrow$

$$\boxed{f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}} = \sum_{j=1}^d\left(f_j(x_j) + \frac{1}{2\lambda}|x_j-y_j|^2\right) \quad (15)}$$

Así tenemos que uniendo (14) y (15) llegamos a que:

$$Prox_{\lambda f}(x) = argmin_{y\in\mathbb{R}^d}\left[f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right] = argmin_{y\in\mathbb{R}^d}\left[f(y) + \frac{1}{2\lambda}\|x-y\|^2_{\mathbf{E}}\right]$$

$$= argmin_{y\in\mathbb{R}^d}\left[\sum_{j=1}^d\left(f_j(x_j) + \frac{1}{2\lambda}|x_j-y_j|^2\right)\right] = argmin_{y_1,\dots,y_d\in\mathbb{R}}\left[\sum_{j=1}^d\left(f_j(x_j) + \frac{1}{2\lambda}|x_j-y_j|^2\right)\right]$$

$$\overset{Problema\ separable}{=} \left(argmin_{y_1\in\mathbb{R}}\left[f_1(x_1) + \frac{1}{2\lambda}|x_1-y_1|^2\right], \dots, argmin_{y_d\in\mathbb{R}}\left[f_d(x_d) + \frac{1}{2\lambda}|x_d-y_d|^2\right]\right)$$

$$\overset{Por\ definicion\ de\ Prox_{\lambda(f_j)}(x_j)\ \forall j\in\{1,\dots,d\}}{=} \left(Prox_{\lambda(f_1)}(x_1), \dots, Prox_{\lambda(f_d)}(x_d)\right)$$

$\Rightarrow$

$$\boxed{Prox_{\lambda f}(x) = \left(Prox_{\lambda(f_1)}(x_1), \dots, Prox_{\lambda(f_d)}(x_d)\right) \quad (16)}$$

que es justamente lo que nos estaban pidiendo demostrar.

Solución d):

Notemos que $f(x) = \|x\|_1 = \sum_{j=1}^{d}|x_j| = \sum_{j=1}^{d} f_j(x_j)$ con $f_j(u) = |u|\ \forall j \in \{1, ..., d\}$.

Ahora por el resultado obtenido en $P2c)$ tenemos que:

$$f_\lambda(x) = \sum_{i=1}^{d}(f_j)_\lambda(x_j)$$

Así lo que haremos será calcular $(f_j)_\lambda(x_j)$ para un $j \in \{1, ..., d\}$ arbitrario.

$$(f_j)_\lambda(x_j) = \inf_{v\in\mathbb{R}} \left[f(v) + \frac{1}{2\lambda}\cdot|x_j - v|^2\right] = \inf_{v\in\mathbb{R}} \left[|v| + \frac{1}{2\lambda}\cdot|x_j - v|^2\right] = \inf_{v\in\mathbb{R}} \left[|v| + \frac{1}{2\lambda}\cdot(x_j - v)^2\right]$$

Lo que haremos, será hacer el cambio de variable $v = t\cdot sign(v)$ con $t \geq 0$.

De esta forma:

$$|v| + \frac{1}{2\lambda}\cdot(x_j - v)^2 = |t\cdot sign(v)| + \frac{1}{2\lambda}\cdot(x_j - t\cdot sign(v))^2 = |t|\cdot|sign(v)| + \frac{1}{2\lambda}\cdot(x_j - t\cdot sign(v))^2$$

$$= t + \frac{1}{2\lambda}\cdot(x_j - t\cdot sign(v))^2 = t + \frac{1}{2\lambda}\left([x_j]^2 - 2t\cdot x_j\cdot sign(v) + t^2[sign(v)]^2\right)$$

$$= t + \frac{1}{2\lambda}\left({x_j}^2 - 2t\cdot x_j\cdot sign(v) + t^2\right)$$

Ahora notemos que como queremos minimizar, entonces para determinar el valor de $sign(v)$ lo que tenemos que hacer es intentar que la expresión anterior sea lo menor posible y como $\lambda > 0$, entonces lo que hay que buscar es que $-2t\cdot x_j\cdot sign(v)$ sea lo menor posible, lo cual es equivalente con que $t\cdot x_j\cdot sign(v)$ sea lo mayor posible. Y como $t \geq 0$, entonces lo que estamos buscando es que $x_j\cdot sign(v)$ sea lo mayor posible.

De este modo, tenemos que $sign(v) = sign(x_j)$ porque así $x_j\cdot sign(v) = x_j\cdot sign(x_j) = |x_j|$.

Así tenemos que la expresión a minimizar es función de t y tiene como restricción que $t \geq 0$, además de que su expresión esta dada por:

$$h(t) = t + \frac{1}{2\lambda}\left({x_j}^2 - 2t\cdot|x_j| + t^2\right)$$

De esta forma:

$$h'(t) = 1 + \frac{1}{2\lambda}\left(-2|x_j| + 2t\right) = 1 + \frac{1}{\lambda}\left(t - |x_j|\right) \;\Rightarrow\; h''(t) = \frac{1}{\lambda} > 0$$

Ahora notemos que como $h''(t) > 0$ entonces h es convexa. De este modo, el mínimo se alcanza en un punto t' tal que $h'(t') = 0$, siempre y cuando el t' sea mayor o igual a cero, de lo contrario el mínimo se alcanzaría en $t' = 0$ (basta con ver el dibujo de una parábola convexa y la ubicación de su mínimo).

$$h(t') = 0 \;\Rightarrow\; 1 + \frac{1}{\lambda}\left(t' - |x_j|\right) = 0 \;\Rightarrow\; \frac{1}{\lambda}\left(t' - |x_j|\right) = -1 \;\Rightarrow\; \frac{1}{\lambda}\left(|x_j| - t'\right) = 1 \;\Rightarrow\; |x_j| - t' = \lambda$$

$$\Rightarrow t' = |x_j| - \lambda$$

Así notemos que si $t' \geq 0$ entonces $|x_j| - \lambda \geq 0 \;\Rightarrow \boxed{|x_j| \geq \lambda}$, entonces el mínimo esta dado por:

$$h(t') = t' + \frac{1}{2\lambda}\left({x_j}^2 - 2t'\cdot|x_j| + [t']^2\right) = |x_j| - \lambda + \frac{1}{2\lambda}\left({x_j}^2 - 2(|x_j| - \lambda)\cdot|x_j| + [|x_j| - \lambda]^2\right)$$

$$= |x_j| - \lambda + \frac{1}{2\lambda}\left({x_j}^2 - 2|x_j|\cdot|x_j| + 2\lambda\cdot|x_j| + |x_j|^2 + \lambda^2 - 2\lambda|x_j|\right)$$

$$= |x_j| - \lambda + \frac{1}{2\lambda}\left({x_j}^2 - 2{x_j}^2 + 2\lambda\cdot|x_j| + {x_j}^2 + \lambda^2 - 2\lambda|x_j|\right)$$

$$= |x_j| - \lambda + \frac{1}{2\lambda}\left(\lambda^2\right) = |x_j| - \lambda + \frac{\lambda}{2} = |x_j| - \frac{\lambda}{2}$$

Ahora recordando que si $|x_j| \geq \lambda$ entonces esto implica que $x_j \in (-\infty, -\lambda] \cup [\lambda, +\infty)$

Así tenemos que:

$$\boxed{\text{Si } x_j \in (-\infty, -\lambda] \cup [\lambda, +\infty) \;\text{ entonces el minimo es } |x_j| - \frac{\lambda}{2} \quad (17)}$$

Por otro lado, tenemos que si $t' < 0$, entonces $|x_j| - \lambda < 0 \;\Rightarrow |x_j| < \lambda$ y el mínimo se da en $t = 0$ y es:

$$h(0) = 0 + \frac{1}{2\lambda}\left({x_j}^2 - 2\cdot 0\cdot|x_j| + 0^2\right) = \frac{1}{2\lambda}\cdot{x_j}^2 = \frac{x_j^2}{2\lambda}$$

Y como sabemos que si $t' < 0$ entonces $|x_j| < \lambda$ y esto implica que $x_j \in (-\lambda, \lambda)$.

De esta forma:

$$\boxed{\text{Si } x_j \in (-\lambda, \lambda) \;\text{ entonces el minimo es } \frac{x_j^2}{2\lambda} \quad (18)}$$

Así tenemos que uniendo (17) y (18) llegamos a que:

$$\boxed{(f_j)_\lambda(x_j) = \begin{cases} |x_j| - \frac{\lambda}{2} & \text{si } x_j \in (-\infty, -\lambda] \cup [\lambda, +\infty) \\ \\ \dfrac{{x_j}^2}{2\lambda} & \text{si } x_j \in (-\lambda, \lambda) \end{cases}}$$

De esta forma tenemos que $f_\lambda(x)$ esta dada por $\sum_{j=1}^{d}(f_j)_\lambda(x_j)$ donde $(f_j)_\lambda(x_j)$ esta dada por la expresión anterior $\forall j \in \{1, ..., d\}$.

Pregunta 3

Parte I:

Primero que todo notemos que para esta sección debemos utilizar el gradiente de $f_\lambda(x)$ que se obtuvo a partir de $f(x) = \|x\|_1$ y para esto usaremos el resultado de $P3d)$. Así tenemos que:

$$\frac{\partial f_\lambda(x)}{\partial x_k} = \frac{\partial}{\partial x_k}\left(\sum_{j=1}^{d}(f_j)_\lambda(x_j)\right) = \sum_{j=1}^{d}\left(\frac{\partial}{\partial x_k}\left[(f_j)_\lambda(x_j)\right]\right) = \sum_{j=1:j\neq k}^{d}\left(\frac{\partial}{\partial x_k}\left[(f_j)_\lambda(x_j)\right]\right)+\frac{\partial}{\partial x_k}\left[(f_k)_\lambda(x_k)\right]$$

$$= \frac{\partial}{\partial x_k}\left[(f_k)_\lambda(x_k)\right]$$

Así que nos centraremos en calcular $\frac{\partial}{\partial x_k}\left[(f_k)_\lambda(x_k)\right]$ usando el resultado de $P3d)$.

Por la expresión de $P3d)$ tenemos que:

$$\frac{\partial}{\partial x_k}\left[(f_k)_\lambda(x_k)\right] = \begin{cases} sign(x_k) & \text{si } x_j \in (-\infty,-\lambda)\cup(\lambda,+\infty) \\ \\ \dfrac{x_k}{\lambda} & \text{si } x_j \in (-\lambda,\lambda) \end{cases}$$

Solo nos falta determinar la derivada parcial en $x_k = \lambda$ y en $x_k = -\lambda$ y lo haremos por definición.

Primero calcularemos la derivada parcial en $x_k = \lambda$ y para esto calcularemos los límites por la derecha e izquierda cuando h tiende a 0 de $\frac{(f_k)_\lambda(\lambda+h)-(f_k)_\lambda(\lambda)}{h}$.

$$\lim_{h\to 0^+}\frac{(f_k)_\lambda(\lambda+h)-(f_k)_\lambda(\lambda)}{h} = \lim_{h\to 0^+}\frac{\left(|\lambda+h|-\frac{\lambda}{2}\right)-\left(|\lambda|-\frac{\lambda}{2}\right)}{h} = \lim_{h\to 0^+}\frac{|\lambda+h|-\frac{\lambda}{2}-|\lambda|+\frac{\lambda}{2}}{h}$$

$$= \lim_{h\to 0^+}\frac{|\lambda+h|-|\lambda|}{h} = \lim_{h\to 0^+}\frac{\lambda+h-\lambda}{h} = \lim_{h\to 0^+}\frac{h}{h} = 1$$

$\Rightarrow$

$$\boxed{\lim_{h\to 0^+}\frac{(f_k)_\lambda(\lambda+h)-(f_k)_\lambda(\lambda)}{h} = 1 \quad (1)}$$

$$\lim_{h\to 0^-}\frac{(f_k)_\lambda(\lambda+h)-(f_k)_\lambda(\lambda)}{h} = \lim_{h\to 0^-}\frac{\left(\frac{(\lambda+h)^2}{2\lambda}\right)-\left(|\lambda|-\frac{\lambda}{2}\right)}{h} = \lim_{h\to 0^-}\frac{\left(\frac{\lambda^2+2h\lambda+h^2}{2\lambda}\right)-\left(\lambda-\frac{\lambda}{2}\right)}{h}$$

$$= \lim_{h\to 0^-}\frac{\frac{\lambda}{2}+h+\frac{h^2}{2\lambda}-\frac{\lambda}{2}}{h} = \lim_{h\to 0^-}\frac{h+\frac{h^2}{2\lambda}}{h} = \lim_{h\to 0^-}\left(1+\frac{h}{2\lambda}\right) = 1$$

$\Rightarrow$

$$\boxed{\lim_{h\to 0^-} \frac{(f_k)_\lambda(\lambda+h)-(f_k)_\lambda(\lambda)}{h} = 1 \quad (2)}$$

Usando (1) y (2) tenemos que existe la derivada en $x_k = \lambda$ y está dada por:

$$\boxed{\lim_{h\to 0} \frac{(f_k)_\lambda(\lambda+h)-(f_k)_\lambda(\lambda)}{h} = 1 \quad (3)}$$

Ahora calcularemos la derivada parcial en $x_k = -\lambda$ y para esto calcularemos los límites por la derecha e izquierda cuando h tiende a 0 de $\frac{(f_k)_\lambda(-\lambda+h)-(f_k)_\lambda(-\lambda)}{h}$.

$$\lim_{h\to 0^+} \frac{(f_k)_\lambda(-\lambda+h)-(f_k)_\lambda(-\lambda)}{h} = \lim_{h\to 0^+} \frac{\left(\frac{(-\lambda+h)^2}{2\lambda}\right)-\left(|-\lambda|-\frac{\lambda}{2}\right)}{h} = \lim_{h\to 0^+} \frac{\left(\frac{\lambda^2-2\cdot\lambda\cdot h+h^2}{2\lambda}\right)-\left(\lambda-\frac{\lambda}{2}\right)}{h}$$

$$= \lim_{h\to 0^+} \frac{\left(\frac{\lambda}{2}-h+\frac{h^2}{2\lambda}\right)-\frac{\lambda}{2}}{h} = \lim_{h\to 0^+} \frac{-h+\frac{h^2}{2\lambda}}{h} = \lim_{h\to 0^+} \left(-1+\frac{h}{2\lambda}\right) = -1$$

$\Rightarrow$

$$\boxed{\lim_{h\to 0^+} \frac{(f_k)_\lambda(-\lambda+h)-(f_k)_\lambda(-\lambda)}{h} = -1 \quad (4)}$$

$$\lim_{h\to 0^-} \frac{(f_k)_\lambda(-\lambda+h)-(f_k)_\lambda(-\lambda)}{h} = \lim_{h\to 0^-} \frac{\left(|-\lambda+h|-\frac{\lambda}{2}\right)-\left(|-\lambda|-\frac{\lambda}{2}\right)}{h}$$

$$= \lim_{h\to 0^-} \frac{\left(\lambda-h-\frac{\lambda}{2}\right)-\left(\lambda-\frac{\lambda}{2}\right)}{h} = \lim_{h\to 0^-} \frac{\left(\frac{\lambda}{2}-h\right)-\left(\frac{\lambda}{2}\right)}{h} = \lim_{h\to 0^-} \frac{-h}{h} = -1$$

$\Rightarrow$

$$\boxed{\lim_{h\to 0^-} \frac{(f_k)_\lambda(-\lambda+h)-(f_k)_\lambda(-\lambda)}{h} = -1 \quad (5)}$$

Usando (4) y (5) tenemos que existe la derivada en $x_k = -\lambda$ y está dada por:

$$\boxed{\lim_{h\to 0} \frac{(f_k)_\lambda(-\lambda+h)-(f_k)_\lambda(-\lambda)}{h} = -1 \quad (6)}$$

De esta forma, uniendo (3) y (6) junto a lo que ya habíamos calculado, tenemos que para calcular el gradiente debemos calcular la derivada parcial en cada variable y estas están dadas por:

$$\boxed{[\nabla f_\lambda(x)]_k = \frac{\partial}{\partial x_k}\left[(f_k)_\lambda(x_k)\right] = \begin{cases} sign(x_k) & \text{si } x_j \in (-\infty, -\lambda] \cup [\lambda, +\infty) \\ \\ \dfrac{x_k}{\lambda} & \text{si } x_j \in (-\lambda, \lambda) \end{cases}}$$

Por otra parte tenemos que para el método del gradiente proyectado lo que tenemos que realizar en uno de los pasos es una proyección sobre el conjunto $C = \{x \in \mathbb{R}^d : Ax = b\}$ y para esto debemos resolver el siguiente problema:

$$\Pi_C(x) = argmin_{v \in \mathbb{R}^d}\{\|v - x\|^2_{\mathbf{E}} : Av = b\}$$

donde $A \in \mathbb{R}^{m \times d}$ y $b \in \mathbb{R}^m$.

Para esto usaremos multiplicadores de lagrange.

Sea $u \in \mathbb{R}^m$ entonces el lagrangeano esta dado por:

$$L(v, u) = \|v - x\|^2_{\mathbf{E}} + u^T(Av - b) = v^T v - 2x^T v + x^T x + u^T Av - u^T b$$

De esta forma:

$$\nabla_v L(v, u) = 2v - 2x + A^T u$$

Y como queremos que $\nabla_v L(v, u) = 0$ entonces tenemos que:

$$2v - 2x + A^T u = 0 \;\Rightarrow\; 2v = 2x - A^T u \;\Rightarrow\; \boxed{v = x - \frac{A^T u}{2} \quad (7)}$$

Pero además se debe cumplir que $Av = b$ y utilizaremos esta igualdad reemplazando en ella el resultado de (7).

De esta forma:

$$A\left(x - \frac{A^T u}{2}\right) = b \;\Rightarrow\; Ax - \frac{AA^T u}{2} = b \;\Rightarrow\; Ax - b = \frac{AA^T u}{2} \;\Rightarrow\; 2(Ax - b) = AA^T u$$

$$\Rightarrow \boxed{u = 2(AA^T)^{-1}(Ax - b) \quad (8)}$$

Reemplazando (8) en (7) llegamos a que:

$$v = x - \frac{A^T[2(AA^T)^{-1}(Ax - b)]}{2} = x - A^T(AA^T)^{-1}(Ax - b)$$

Y como v es el argmin, entonces:

$$\boxed{\Pi_C(x) = x - A^T(AA^T)^{-1}(Ax - b) \quad (9)}$$

Además de esto notemos que para el método del gradiente proyectado un paso que siempre cumple el lema de descenso y que además es el paso optimo es $\frac{1}{\mu}$ donde μ es la constante de suavidad de la función a considerar. Y como para la regularización sabemos que la constante de suavidad es $\frac{1}{lambda}$, entonces el paso para aplicar gradiente proyectado sobre la regularización será: λ.

De esta forma se implementará el método de la siguiente forma:

Método del gradiente proyectado

0. Dado $x_0 \in E$, $T \in \mathbb{N}$ y $\lambda > 0$.

1. $x_0 = \Pi_C(x_0) = x_0 - A^T(AA^T)^{-1}(Ax_0 - b)$

2. *For* $t = 0, ..., T-1$:

$$z^t = x^t - \lambda \cdot \nabla f_\lambda(x^t)$$

$$x^{t+1} = \Pi_C(z^t) = z^t - A^T(AA^T)^{-1}(Az^t - b)$$

3. return x^T

También tenemos que implementar el método de Nesterov, donde la función objetivo es:

$$g(x) = f_\lambda(x) + \frac{\mu}{2} \cdot \|Ax - b\|_{\mathbf{E}}^2$$

Y para esto debemos calcular el gradiente de $g(x)$ donde el gradiente de $f_\lambda(x)$ ya lo sabemos, por ende solo nos falta calcular el gradiente de $\frac{\mu}{2} \cdot \|Ax - b\|_{\mathbf{E}}^2$.

$$\frac{\mu}{2} \cdot \|Ax-b\|_{\mathbf{E}}^2 = \frac{\mu}{2}(Ax-b)^T(Ax-b) = \frac{\mu}{2}(x^TA^T - b^T)(Ax-b) = \frac{\mu}{2}(x^TA^TAx - b^TAx - x^TA^Tb + b^Tb)$$

$$= \frac{\mu}{2}(x^TA^TAx - 2b^TAx + b^Tb) = \frac{\mu}{2}(x^TA^TAx + [-2A^Tb]^Tx + b^Tb)$$

$$\Rightarrow \nabla_x\left(\frac{\mu}{2} \cdot \|Ax - b\|_{\mathbf{E}}^2\right) = \frac{\mu}{2} \cdot (2A^TAx - 2A^Tb) = \mu \cdot A^T(Ax - b)$$

De esta forma:

$$\boxed{\nabla g(x) = \nabla f_\lambda(x) + \mu \cdot A^T(Ax - b) \quad (10)}$$

Así utilizando este resultado implementaremos el método de Nesterov.

Ahora notemos que la función que queremos minimizar es $f_\lambda(x) + \frac{\mu}{2}\|Ax - b\|_2$ y para determinar el η_{optimo} debemos saber la constante de suavidad de esta función, pues el η_{optimo} es el inverso de esta constante de suavidad.

Considerando la función objetivo, tenemos que la contante de suavidad de esta función es:

$$\frac{1}{\lambda} + \mu \cdot \lambda_{max}(A^T A)$$

Y como sabemos $\lambda_{max}(A^T A) \geq 1$ pues $\lambda_{max}(A^T A)$ es la norma 2 de A y las normas matriciales son siempre mayor o igual a 1.

De este modo:

$$\frac{1}{\lambda} + \mu \cdot \lambda_{max}(A^T A) \geq \frac{1}{\lambda} + \mu$$

$\Rightarrow$

$$\frac{1}{\frac{1}{\lambda} + \mu \cdot \lambda_{max}(A^T A)} \leq \frac{1}{\frac{1}{\lambda} + \mu}$$

$\Rightarrow$

$$\eta_{optimo} \leq \frac{1}{\frac{1}{\lambda} + \mu}$$

Y para nuestro caso en particular tomaremos $\eta = \frac{1}{\frac{1}{\lambda}+\mu}$.

Así tenemos que el método de Nesterov que implementaremos es el siguiente.

Método de Nesterov

1. $x^0 = y^0 \in E$, $\eta = \frac{1}{\frac{1}{\lambda}+\mu}$.

2. For $t = 0, ..., T-1$:

$$x^{t+1} = y^t - \eta \cdot [\nabla f_\lambda(y^t) + \mu \cdot A^T(Ay^t - b)]$$

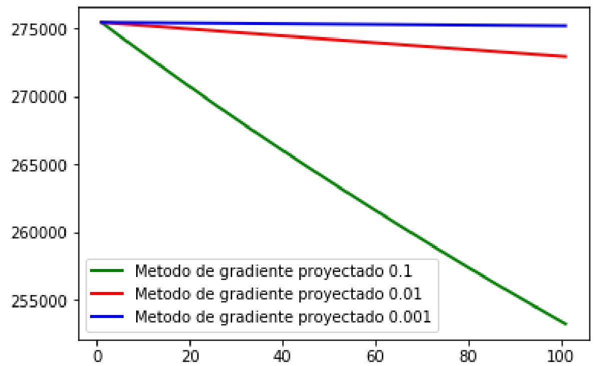
$$y^{t+1} = x^{t+1} + \left(\tfrac{t}{t+3}\right) \cdot (x^{t+1} - x^t)$$

3. return y^T

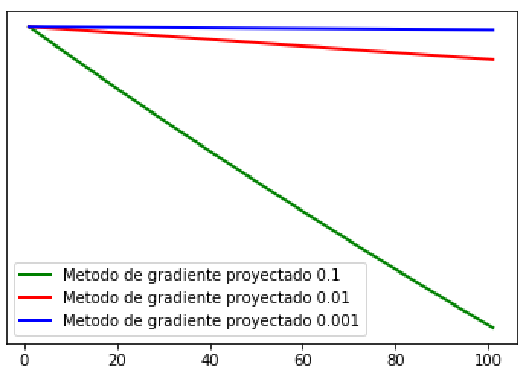
Resultados computacionales:

Lo primero que hicimos en la parte de resultados computacionales fue aplicar el método del gradiente proyectado a la regularizada, es decir f_λ (para $\lambda = 10^{-1}, 10^{-2}, 10^{-3}$) obteniéndose los puntos $x^0, x^1, ..., x^T$. Y para cada uno de estos puntos se gráfico $f(x^t) = \|x^t\|_1$ (Se realizaron 100 iteraciones).

El gráfico en escala normal de $\|\cdot\|_1$ en función de las iteraciones es:



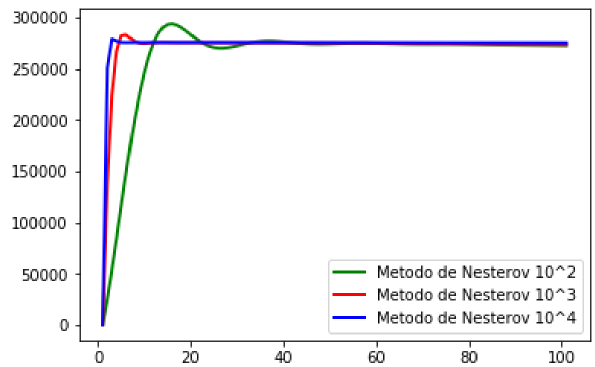
El gráfico en escala logaritmica de $\|\cdot\|_1$ en función de las iteraciones es:



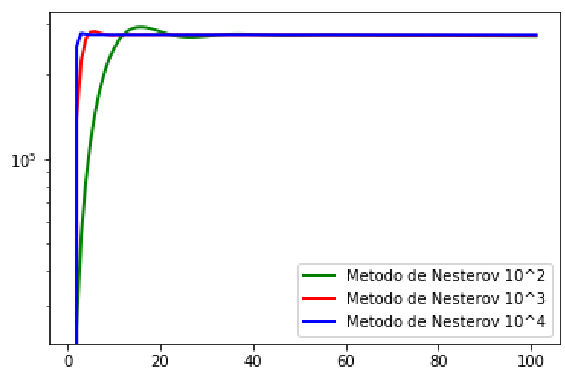
<u>Comentarios:</u> Como se observa a medida que disminuye al valor de λ, aumenta la norma 1 de los valores $x_0, x_1, ..., x^T$ (ver gráficos). Por otro lado, notemos que como vimos en la pregunta 1a), cuando $\lambda \to 0^+$ entonces la regularizada converge a $f(x)$, por lo tanto, considerando esto podemos decir que lo ideal sería elegir un λ lo menor posible, sin embargo, hay un tradeoff en el sentido de que el paso en el método del gradiente proyectado es λ, por ende si disminuimos el valor de λ, los pasos del algoritmo son cada vez menores. Finalmente elegimos utilizar el valor de $\lambda = 10^{-3}$.

Luego, utilizamos este $\lambda = 10^{-3}$ y ejecutamos el método de Nesterov para valores de $\mu = 10^2, 10^3, 10^4$ obteniendo los puntos $y^0, y^1, ..., y^T$ y para estos puntos se graficaron los valores de la norma 1 para estos puntos obteniéndose los siguientes gráficos (Se realizaron 100 iteraciones).

El gráfico en escala normal de $\|\cdot\|_1$ en función de las iteraciones es:



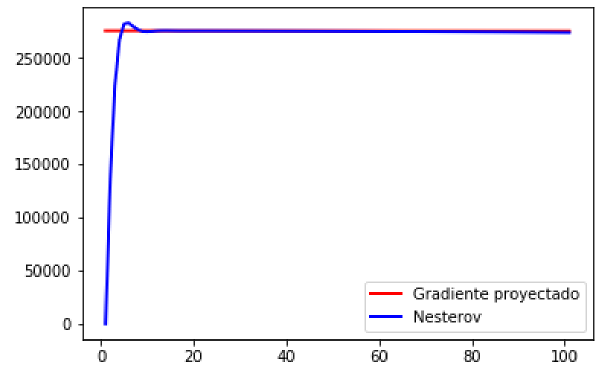
El gráfico en escala logarítmica de $\|\cdot\|_1$ en función de las iteraciones es:



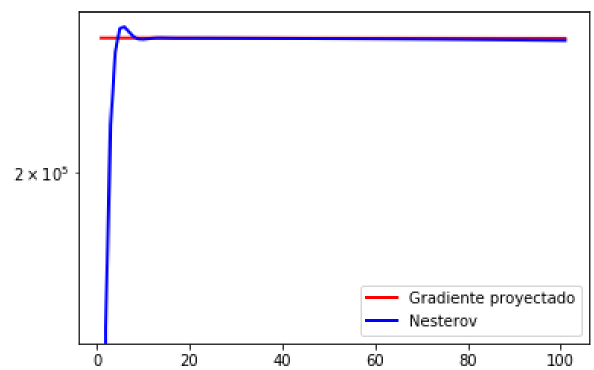
Comentarios: Recordemos que para nosotros lo ideal es que u sea lo mayor posible. Como se observa en ambos gráficos, la similitud entre los gráficos del método de Nesterov para $\mu = 10^4$ y $\mu = 10^3$ es mucho mas evidente que la similitud entre los métodos de Nesterov para $\mu = 10^2$ y $\mu = 10^3$. Pero al igual que en el caso anterior, hay un tradeoff, pues a pesar de que queremos el valor de μ lo más grande posible, a medida que aumenta el valor de μ, el valor de $\eta_{optimo} = \frac{1}{\frac{1}{\lambda}+\mu}$ disminuye. Finalmente elegimos utilizar $\mu = 10^3$.

Y finalmente graficaremos el método del gradiente proyectado y el método de Nesterov para 100 iteraciones con $\lambda = 10^{-3}$ y $\mu = 10^{3}$.

El gráfico de $\|\cdot\|_1$ en función de las iteraciones en escala normal es:



El gráfico de $\|\cdot\|_1$ en función de las iteraciones en escala logarítmica es:



<u>Comentarios:</u> Notemos que tanto el método de Nesterov como el método del gradiente proyectado se comportan muy similares a medida que avanzan las iteraciones y además convergen a valores similares. Sin embargo, el método de Nesterov demora mucho menos tiempo en ejecutarse que el método del gradiente proyectado y esto se puede explicar porque para el método de Nesterov solo necesitamos calcular gradiente, mientras que el método del gradiente proyectado, necesitamos calcular gradientes además de la proyección en donde debemos multiplicar e invertir matrices lo cual toma mucho tiempo.

Parte II:

El método que nosotros implementaremos es una combinación de Nesterov y de Gradiente proyectado que primero realizar 9 iteraciones de Nesterov, luego una iteración de gradiente proyectado, y luego nuevamente 9 iteraciones de gradiente proyectado y así. Además de eso, una vez que terminemos las iteraciones (realizaremos 100), todas las componentes del vector solución que tienen un valor absoluto menor a 1 los haremos 0, para que así realmente nos quede un vector sparse.

De este modo, el método sería de la forma (al método le puse nombre de combinación):

def combinación:

$x_0 \in E$

$x = x_0 \in E$

for $t_1 = 1, ..., 10$:

for $t_2 = 1, ..., 9$:

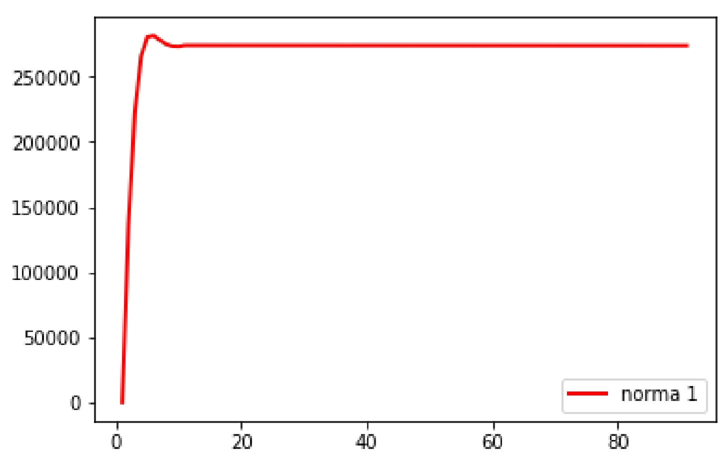
$x = Nesterov(x)$

$x = GradienteProyectado(x)$

$procesamiento(y)$

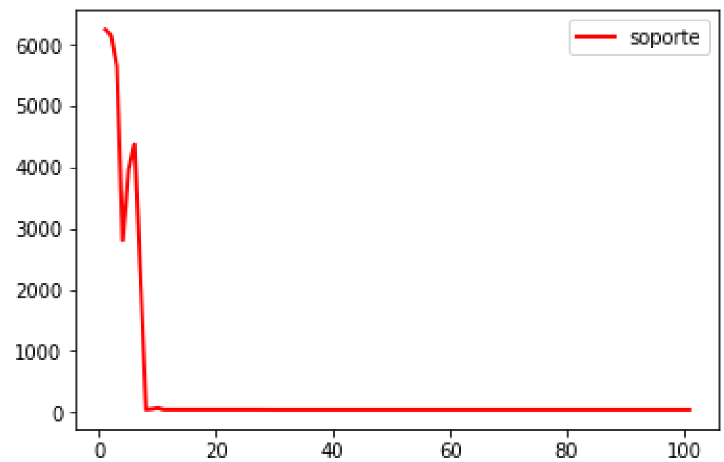
return y

De este modo, obtenemos que la norma 1 en cada iteración, las cuales se pueden ver en el siguiente gráfico:



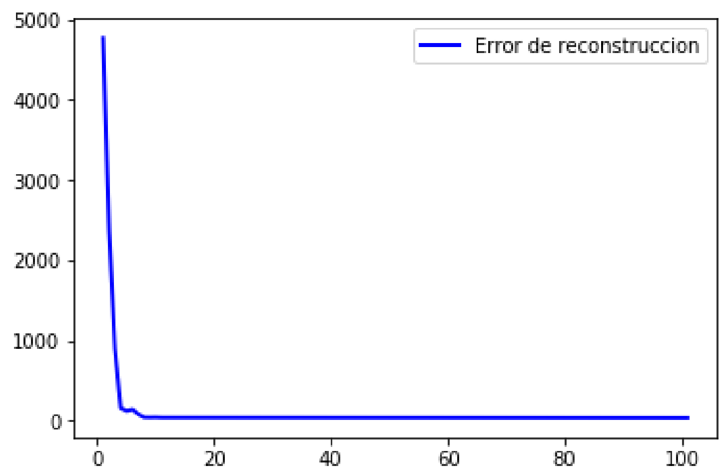
Como se observa, la norma 1 para este algoritmo se comporta muy similar a la de los otros algoritmos.

Por otro lado, comenzaremos a mostrar los gráficos que nos piden.



Para este gráfico, notamos que el la diferencia en norma 0 comienza en un inicio en valores mayores a 6,000 y considerando que el mayor valor posible en norma 0 será 6400 pues el vector es de dimensión 6,400, entonces podemos decir que comienza mal, pero al final termina convergiendo a valores cercanos a 0, por lo tanto, podemos decir que el método propuesto converge.

ii. Error de reconstrucción:



Como se observa, el método converge también cuando estamos tomando norma 2, y según ambos gráficos se ve que el método converge en aproximadamente 15 iteraciones, por lo tanto, es justamente para esta cantidad de iteraciones es que veremos cuanto se demora el algoritmo.

iii. Velocidad de convergencia:

$$1{,}5780190890000085 \text{ segundos}$$

Esta medida del tiempo que se demoró en iterar, lo consideraremos como la velocidad de convergencia.