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

Pregunta Bonus I1

Pregunta

Lo primero que debemos tener en cuenta es la siguiente desigualdad:

$$\boxed{\|u - \Pi_{\mathcal{X}}(u)\|_2^2 \leq \|u - y\|_2^2 \quad \forall y \in E \quad (1)}$$

Por definición de $\Pi_{\mathcal{X}}$ (todo esto para el $\mathcal{X}$ que se definió en la prueba).

Usando (1) en $u = x + h$, $y = \Pi_{\mathcal{X}}(x)$, entonces:

$$\boxed{\|x + h - \Pi_{\mathcal{X}}(x+h)\|_2^2 \leq \|x + h - \Pi_{\mathcal{X}}(x)\|_2^2 \quad (2)}$$

De este modo:

$$f(x+h) = \frac{1}{2}\|x + h - \Pi_{\mathcal{X}}(x+h)\|_2^2 \leq \frac{1}{2}\|x + h - \Pi_{\mathcal{X}}(x)\|_2^2 = \frac{1}{2}\|[x - \Pi_{\mathcal{X}}(x)] + h\|_2^2$$

$$= \frac{1}{2}\left(\|x - \Pi_{\mathcal{X}}(x)\|_2^2 + 2 < x - \Pi_{\mathcal{X}}(x), h > + \|h\|_2^2\right) = \frac{1}{2}\|x - \Pi_{\mathcal{X}}(x)\|_2^2 + < x - \Pi_{\mathcal{X}}(x), h > + \frac{1}{2}\|h\|_2^2$$

$$= f(x) + < x - \Pi_{\mathcal{X}}(x), h > + \frac{1}{2}\|h\|_2^2$$

$\Rightarrow$

$$f(x+h) \leq f(x) + < x - \Pi_{\mathcal{X}}(x), h > + \frac{1}{2}\|h\|_2^2 \quad (3)$$

Ahora notemos que usando que $\nabla f(x) = x - \Pi_{\mathcal{X}}(x)$ y reemplazando en (3) tenemos que:

$$f(x+h) \leq f(x) + < \nabla f(x), h > + \frac{1}{2}\|h\|_2^2$$

$\Rightarrow$

$$f(x+h) - f(x) - < \nabla f(x), h > \leq \frac{1}{2}\|h\|_2^2$$

$\Rightarrow$

$$\frac{f(x+h) - f(x) - < \nabla f(x), h >}{\|h\|_2} \leq \frac{1}{2}\|h\|_2$$

Ahora si aplicamos el limite de cuando $h \to 0$, entonces tenemos que:

$$\lim_{h\to 0} \frac{f(x+h)-f(x)-<\nabla f(x),h>}{\|h\|_2} \leq \lim_{h\to 0} \frac{1}{2}\|h\|_2 = 0$$

De este modo, llegamos a que:

$$\boxed{\lim_{h\to 0} \frac{f(x+h)-f(x)-<\nabla f(x),h>}{\|h\|_2} \leq 0 \quad (4)}$$

Ahora usando (1) con $u = x$, $y = \Pi_{\mathcal{X}}(x+h)$, llegamos a que:

$$\|x-\Pi_{\mathcal{X}}(x)\|_2^2 \leq \|x-\Pi_{\mathcal{X}}(x+h)\|_2^2$$

$\Rightarrow$

$$\boxed{-\frac{1}{2}\cdot\|x-\Pi_{\mathcal{X}}(x)\|_2^2 \geq -\frac{1}{2}\cdot\|x-\Pi_{\mathcal{X}}(x+h)\|_2^2 \quad (5)}$$

De este modo, tenemos que usando (5) llegamos a que:

$$f(x+h)-f(x) = f(x+h)-\frac{1}{2}\cdot\|x-\Pi_{\mathcal{X}}(x)\|_2^2 \geq f(x+h)-\frac{1}{2}\cdot\|x-\Pi_{\mathcal{X}}(x+h)\|_2^2$$

$$= \frac{1}{2}\cdot\|x+h-\Pi_{\mathcal{X}}(x+h)\|_2^2-\frac{1}{2}\cdot\|x-\Pi_{\mathcal{X}}(x+h)\|_2^2 = \frac{1}{2}\cdot\left(\|[x-\Pi_{\mathcal{X}}(x+h)]+h\|_2^2-\|x-\Pi_{\mathcal{X}}(x+h)\|_2^2\right)$$

$$= \frac{1}{2}\cdot\left(\|[x-\Pi_{\mathcal{X}}(x+h)]\|_2^2+2<x-\Pi_{\mathcal{X}}(x+h),h>+\|h\|_2^2-\|x-\Pi_{\mathcal{X}}(x+h)\|_2^2\right)$$

$$= \frac{1}{2}\cdot\left(2<x-\Pi_{\mathcal{X}}(x+h),h>+\|h\|_2^2\right) = <x-\Pi_{\mathcal{X}}(x+h),h>+\frac{1}{2}\cdot\|h\|_2^2$$

$$= <x-\Pi_{\mathcal{X}}(x)+\Pi_{\mathcal{X}}(x)-\Pi_{\mathcal{X}}(x+h),h>+\frac{1}{2}\cdot\|h\|_2^2$$

$$= <x-\Pi_{\mathcal{X}}(x),h>+<\Pi_{\mathcal{X}}(x)-\Pi_{\mathcal{X}}(x+h),h>+\frac{1}{2}\cdot\|h\|_2^2$$

$\Rightarrow$

$$f(x+h)-f(x) \geq \ <x-\Pi_{\mathcal{X}}(x),h>+<\Pi_{\mathcal{X}}(x)-\Pi_{\mathcal{X}}(x+h),h>+\frac{1}{2}\cdot\|h\|_2^2$$

$\Rightarrow$

$$f(x+h)-f(x)-<x-\Pi_{\mathcal{X}}(x),h> \ \geq \ <\Pi_{\mathcal{X}}(x)-\Pi_{\mathcal{X}}(x+h),h>+\frac{1}{2}\cdot\|h\|_2^2$$

Usando que $x - \Pi_{\mathcal{X}}(x) = \nabla f(x)$ tenemos que:

$$f(x+h) - f(x) - < \nabla f(x), h > \ \geq \ < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), h > + \frac{1}{2} \cdot \|h\|_2^2$$

$\Rightarrow$

$$\frac{f(x+h) - f(x) - < \nabla f(x), h >}{\|h\|_2} \ \geq \ \frac{1}{\|h\|_2} < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), h > + \frac{1}{2} \cdot \|h\|_2$$

Ahora aplicamos el limite cuando $h \to 0$ a ambos lados y tenemos que:

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - < \nabla f(x), h >}{\|h\|_2} \ \geq \ \lim_{h \to 0} \left(\frac{1}{\|h\|_2} < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), h > + \frac{1}{2} \cdot \|h\|_2 \right)$$

$$= \lim_{h \to 0} \left(\frac{1}{\|h\|_2} < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), h > \right) + \lim_{h \to 0} \frac{1}{2} \cdot \|h\|_2 = \lim_{h \to 0} \left(\frac{1}{\|h\|_2} < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), h > \right)$$

$$= \lim_{h \to 0} < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} >$$

$\Rightarrow$

$$\boxed{\lim_{h \to 0} \frac{f(x+h) - f(x) - < \nabla f(x), h >}{\|h\|_2} \geq \lim_{h \to 0} < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > \qquad (6)}$$

Ahora notemos que:

$$\boxed{0 \leq | < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | \qquad (7)}$$

Y usando $Cauchy - Schwarz$ tenemos que:

$$| < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | \ \leq \ \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h)\| \cdot \|\frac{h}{\|h\|_2}\|_2$$

$$= \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h)\| \cdot \frac{\|h\|_2}{\|h\|_2} = \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h)\|$$

$\Rightarrow$

$$\boxed{| < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | \ \leq \ \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h)\| \qquad (8)}$$

Usando que las proyecciones son Lipschitz con $L = 1$ llegamos a que:

$$\boxed{\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h)\| \leq \|x - (x+h)\|_2 = \|-h\|_2 = \|h\|_2 \qquad (9)}$$

Ahora usando (9) en (8) tenemos que:

$$| < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | \ \leq \ \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h)\| \leq \|h\|_2$$

$\Rightarrow$

$$\boxed{| < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | \ \leq \ \|h\|_2 \quad (10)}$$

Ahora usando (7) y (10) tenemos que:

$$0 \leq | < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | \leq \|h\|_2$$

Aplicando limite a ambos lados de cuando $h \to 0$ tenemos que:

$$\lim_{h\to 0} 0 \leq \ \lim_{h\to 0} | < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | \ \leq \lim_{h\to 0} \|h\|_2$$

$\Rightarrow$

$$0 \leq \ \lim_{h\to 0} | < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | \ \leq 0$$

$\Rightarrow$

$$\boxed{\lim_{h\to 0} | < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | = 0 \quad (11)}$$

Ahora como la función $|\cdot|$ es continua, entonces tenemos que:

$$| \lim_{h\to 0} < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > | = 0$$

$\Rightarrow$

$$\boxed{\lim_{h\to 0} < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} > = 0 \quad (12)}$$

Ahora reemplazando (12) en (6) llegamos a que:

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - < \nabla f(x), h >}{\|h\|_2} \geq \lim_{h\to 0} < \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(x+h), \frac{h}{\|h\|_2} >= 0$$

$\Rightarrow$

$$\boxed{\lim_{h\to 0} \frac{f(x+h) - f(x) - < \nabla f(x), h >}{\|h\|_2} \geq 0 \quad (13)}$$

Usando (13) y (4) tenemos que:

$$0 \leq \lim_{h \to 0} \frac{f(x+h) - f(x) - < \nabla f(x), h >}{\|h\|_2} \leq 0$$

$\Rightarrow$

$$\boxed{\lim_{h \to 0} \frac{f(x+h) - f(x) - < \nabla f(x), h >}{\|h\|_2} = 0 \quad (14)}$$

Y eso implica que f es diferenciable en x. Pero como x es arbitrario, entonces se demostró que f es diferenciable en todo x en E. Y esto es lo mismo a decir que f es diferenciable en E, que es justamente lo que nos pedían demostrar.

Y como para demostrar esto se uso que $\nabla f(x) = x - \Pi_{\mathcal{X}}(x)$, entonces se demostro que efectivamente ese el gradiente.

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

Tarea 2

Pregunta 1

a) Recordemos que la constante de curvatura se define como:

$$C_f = \sup_{z,x\in\mathcal{X}:\lambda\in[0,1],y=(1-\lambda)x+\lambda z} \frac{2}{\lambda^2}[f(y) - f(x) - < \nabla f(x), y - x >]$$

$$= \sup_{z,x\in\mathcal{X}:\lambda\in[0,1]} \frac{2}{\lambda^2}[f([1-\lambda]x + \lambda z) - f(x) - < \nabla f(x), ([1-\lambda]x + \lambda z) - x >]$$

$$= \sup_{z,x\in\mathcal{X}:\lambda\in[0,1]} \frac{2}{\lambda^2}[f([1-\lambda]x + \lambda z) - f(x) - < \nabla f(x), \lambda(z - x) >]$$

$$= \sup_{z,x\in\mathcal{X}:\lambda\in[0,1]} \frac{2}{\lambda^2}[f([1-\lambda]x + \lambda z) - f(x) - \lambda < \nabla f(x), z - x >]$$

Ahora usando que $f(x) = \frac{1}{2} \cdot \|x\|_2^2 = \frac{1}{2} \cdot x^T x$, entonces tenemos que: $\nabla f(x) = x$.

De este modo:

$$f([1-\lambda]x + \lambda z) = \frac{1}{2} \cdot ([1-\lambda]x + \lambda z)^T([1-\lambda]x + \lambda z) = \frac{1}{2} \cdot ([1-\lambda]x^T + \lambda z^T)([1-\lambda]x + \lambda z)$$

$$\frac{1}{2} \cdot ([1-\lambda]^2 x^T x + \lambda[1-\lambda]z^T x + \lambda[1-\lambda]x^T z + \lambda^2 z^T z)$$

Ahora recordando que $z^T x = x^T z$, entonces tenemos que:

$$f([1-\lambda]x+\lambda z) = \frac{1}{2}\cdot([1-\lambda]x+\lambda z)^T([1-\lambda]x+\lambda z) = \frac{1}{2}\cdot([1-\lambda]^2 x^T x+\lambda[1-\lambda]z^T x+\lambda[1-\lambda]x^T z+\lambda^2 z^T z)$$

$$= \frac{1}{2} \cdot ([1-\lambda]^2 x^T x + \lambda[1-\lambda]x^T z + \lambda[1-\lambda]x^T z + \lambda^2 z^T z) = \frac{1}{2} \cdot ([1-\lambda]^2 x^T x + 2\lambda[1-\lambda]x^T z + \lambda^2 z^T z)$$

$$= \frac{1}{2} \cdot ([1 - 2\lambda + \lambda^2]x^T x + [2\lambda - 2\lambda^2]x^T z + \lambda^2 z^T z)$$

$$= \frac{1}{2} \cdot (x^T x - 2\lambda x^T x + \lambda^2 x^T x + 2\lambda x^T z - 2\lambda^2 x^T z + \lambda^2 z^T z)$$

De este modo:

$$\boxed{f([1-\lambda]x + \lambda z) = \frac{1}{2} \cdot (x^T x - 2\lambda x^T x + \lambda^2 x^T x + 2\lambda x^T z - 2\lambda^2 x^T z + \lambda^2 z^T z) \quad (1)}$$

Además tenemos que:

$$\boxed{-f(x) = -\frac{1}{2} x^T x \quad (2)}$$

Recordando que $\nabla f(x) = x$, entonces:

$$-\lambda < \nabla f(x), z - x >= -\lambda < x, z - x >= -\lambda[x^T(z-x)] = -\lambda[x^T z - x^T x] = -\lambda x^T z + \lambda x^T x$$

De este modo:

$$\boxed{-\lambda < \nabla f(x), z - x >= -\lambda x^T z + \lambda x^T x \quad (3)}$$

Usando $(1), (2)$ y (3) llegamos a que:

$$f([1-\lambda]x + \lambda z) - f(x) - \lambda < \nabla f(x), z - x >= \frac{1}{2} \cdot (x^T x - 2\lambda x^T x + \lambda^2 x^T x + 2\lambda x^T z - 2\lambda^2 x^T z + \lambda^2 z^T z)$$

$$-\frac{1}{2} x^T x - \lambda x^T z + \lambda x^T x = \frac{1}{2} \cdot (x^T x - 2\lambda x^T x + \lambda^2 x^T x + 2\lambda x^T z - 2\lambda^2 x^T z + \lambda^2 z^T z) - \frac{1}{2} x^T x - \frac{1}{2}(2\lambda x^T z) + \frac{1}{2}(2\lambda x^T x)$$

$$= \frac{(x^T x - 2\lambda x^T x + \lambda^2 x^T x + 2\lambda x^T z - 2\lambda^2 x^T z + \lambda^2 z^T z - x^T x - 2\lambda x^T z + 2\lambda x^T x)}{2}$$

$$= \frac{(\lambda^2 x^T - 2\lambda^2 x^T z + \lambda^2 z^T z)}{2} = \frac{2}{\lambda^2}(x^T - 2x^T z + z^T z) = \frac{2}{\lambda^2} \|x - z\|_2^2$$

De este modo:

$$f([1-\lambda]x + \lambda z) - f(x) - \lambda < \nabla f(x), z - x >= \frac{2}{\lambda^2} \|x - z\|_2^2$$

$\Rightarrow$

$$\frac{2}{\lambda^2}[f([1-\lambda]x + \lambda z) - f(x) - \lambda < \nabla f(x), z - x >] = \|x - z\|_2^2$$

Notando que $\lambda \in [0, 1]$, entonces:

$\Rightarrow$

$$C_f = \sup_{z,x\in\mathcal{X}} \left(\frac{2}{\lambda^2}[f([1-\lambda]x+\lambda z) - f(x) - \lambda < \nabla f(x), z-x >] \right) = \sup_{z,x\in\mathcal{X}} \|x-z\|_2^2 = [diam_{\|\cdot\|_2}(\mathcal{X})]^2$$

Así finalmente se llego a que:

$$\boxed{C_f = [diam_{\|\cdot\|_2}(\mathcal{X})]^2}$$

Nota: Como $\mathcal{X}$ es compacto, entonces es acotado, por ende $diam_{\|\cdot\|_2}(\mathcal{X})$ es finito y por ende la constante de curvatura C_f es finita.

---

b) Como $\mu \geq 0$ es tal que $f \in \mathcal{F}^1_{\mathcal{X},\|\cdot\|}(\mu)$, entonces:

$$f(y) \leq f(x) + < \nabla f(x), y-x > + \frac{\mu}{2}\|y-x\|^2 \quad \forall x,y \in \mathcal{X}$$

Sean $z,x \in \mathcal{X}$ y $\lambda \in [0,1]$, entonces definimos $y = (1-\lambda)x + \lambda z$ el cual pertenece a $\mathcal{X}$ pues $\mathcal{X}$ es convexo. De este modo:

$$f(y) \leq f(x) + < \nabla f(x), y-x > + \frac{\mu}{2}\|y-x\|^2 = f(x) + < \nabla f(x), y-x > + \frac{\mu}{2}\|[(1-\lambda)x+\lambda z] - x\|^2$$

$$= f(x) + < \nabla f(x), y-x > + \frac{\mu}{2}\|\lambda(z-x)\|^2 = f(x) + < \nabla f(x), y-x > + \frac{\mu}{2}\lambda^2\|z-x\|^2$$

$\Rightarrow$

$$f(y) \leq f(x) + < \nabla f(x), y-x > + \frac{\mu}{2}\lambda^2\|z-x\|^2$$

$\Rightarrow$

$$f(y) - f(x) - < \nabla f(x), y-x > \leq \frac{\mu}{2}\lambda^2\|z-x\|^2$$

$\Rightarrow$

$$\frac{2}{\lambda^2}[f(y) - f(x) - < \nabla f(x), y-x >] \ \leq \mu\|z-x\|^2$$

Como $y = (1-\lambda)x + \lambda z$ y $\lambda \in [0,1]$, entonces:

$$C_f = sup_{x,z\in\mathcal{X}: y=(1-\lambda)x+\lambda z, \lambda\in[0,1]} \ \frac{2}{\lambda^2}[f(y)-f(x)- < \nabla f(x), y-x >] \ \leq sup_{x,z\in\mathcal{X}: y=(1-\lambda)x+\lambda z, \lambda\in[0,1]} \ \mu\cdot\|z-x\|^2$$

$$= sup_{x,z\in\mathcal{X}} \ \mu\cdot\|z-x\|^2 = \mu\cdot sup_{x,z\in\mathcal{X}} \ \|z-x\|^2 = \mu\cdot[diam_{\|\cdot\|}(\mathcal{X})]^2$$

De este modo, llegamos a que:

$$\boxed{C_f \leq \mu \cdot [diam_{\|\cdot\|}(\mathcal{X})]^2}$$

---

c) Recordando que el gradiente condicional nos plantea que $x^{t+1} = x^t + \eta_t(z^t - x^t)$ con $\eta_t \in [0, 1]$.

De este modo:

$$C_f = \sup_{z,x\in\mathcal{X}:\lambda\in[0,1],y=(1-\lambda)x+\lambda z} \frac{2}{\lambda^2}[f(y) - f(x) - < \nabla f(x), y - x >]$$

$$= \sup_{z,x\in\mathcal{X}:\eta_t\in[0,1]} \frac{2}{[\eta_t]^2}[f([1-\eta_t]x + \eta_t z) - f(x) - < \nabla f(x), [(1-\eta_t)x + \eta_t z] - x >]$$

Tomamos $z = z^t$ y $x = x^t$, entonces:

$$\sup_{z,x\in\mathcal{X}:\eta_t\in[0,1]} \frac{2}{[\eta_t]^2}[f([1-\eta_t]x + \eta_t z) - f(x) - < \nabla f(x), [(1-\eta_t)x + \eta_t z] - x >]$$

$$\geq \frac{2}{[\eta_t]^2}[f([1-\eta_t]x^t + \eta_t z^t) - f(x^t) - < \nabla f(x^t), [(1-\eta_t)x^t + \eta_t z^t] - x >]$$

$$= \frac{2}{[\eta_t]^2}[f(x^t + \eta_t[z^t - x^t]) - f(x^t) - < \nabla f(x^t), (x^t + \eta_t[z^t - x^t]) - x^t >]$$

$$= \frac{2}{[\eta_t]^2}[f(x^t + \eta_t[z^t - x^t]) - f(x^t) - \eta_t < \nabla f(x^t), z^t - x^t >]$$

Usando que $x^{t+1} = x^t + \eta_t(z^t - x^t)$, entonces:

$$= \frac{2}{[\eta_t]^2}[f(x^t+\eta_t[z^t-x^t]) - f(x^t) - \eta_t < \nabla f(x^t), z^t - x^t >] = \frac{2}{[\eta_t]^2}[f(x^{t+1}) - f(x^t) - \eta_t < \nabla f(x^t), z^t - x^t >]$$

De esta forma:

$$C_f \geq \frac{2}{[\eta_t]^2}[f(x^{t+1}) - f(x^t) - \eta_t < \nabla f(x^t), z^t - x^t >]$$

$\Rightarrow$

$$\frac{C_f \cdot \eta_t^2}{2} \geq f(x^{t+1}) - f(x^t) - \eta_t < \nabla f(x^t), z^t - x^t >$$

$\Rightarrow$

$$f(x^t) + \eta_t < \nabla f(x^t), z^t - x^t > + \frac{C_f \cdot \eta_t^2}{2} \geq f(x^{t+1})$$

Así finalmente llegamos a que:

$$\boxed{f(x^{t+1}) \leq f(x^t) + \eta_t < \nabla f(x^t), z^t - x^t > + \frac{C_f \cdot \eta_t^2}{2}}$$

Que es justamente lo pedido.

---

d) Usando la desigualdad de c) tenemos que:

$$f(x^{t+1}) \leq f(x^t) + \eta_t < \nabla f(x^t), z^t - x^t > + \frac{C_f \cdot \eta_t^2}{2}$$

$\Rightarrow$

$$f(x^{t+1}) - f(x^t) \leq \eta_t < \nabla f(x^t), z^t - x^t > + \frac{C_f \cdot \eta_t^2}{2} \quad (1)$$

Ahora por como se definió z^t tenemos que:

$$< \nabla f(x^t), z^t - x^t > = < \nabla f(x^t), z^t > - < \nabla f(x^t), x^t > = min_{z \in \mathcal{X}} < \nabla f(x^t), z > - < \nabla f(x^t), x^t >$$

$$\leq < \nabla f(x^t), x^t > - < \nabla f(x^t), x^t > = 0$$

$\Rightarrow$

$$< \nabla f(x^t), z^t - x^t > \leq 0 \quad (2)$$

Usando (2) en (1) tenemos que:

$$f(x^{t+1}) - f(x^t) \leq \eta_t < \nabla f(x^t), z^t - x^t > + \frac{C_f \cdot \eta_t^2}{2} \leq \frac{C_f \cdot \eta_t^2}{2}$$

$\Rightarrow$

$$f(x^{t+1}) - f(x^t) \leq \frac{C_f \cdot \eta_t^2}{2}$$

Ahora aplicamos una sumatoria desde $t = 0$ hasta $t = T - 1$ en ambos lados de la desigualdad:

$$\sum_{t=0}^{T-1} [f(x^{t+1}) - f(x^t)] \leq \sum_{t=0}^{T-1} \frac{C_f \cdot \eta_t^2}{2} = \frac{C_f}{2} \sum_{t=0}^{T-1} \eta_t^2$$

$\Rightarrow$

$$\sum_{t=0}^{T-1} [f(x^{t+1}) - f(x^t)] \leq \frac{C_f}{2} \sum_{t=0}^{T-1} \eta_t^2 \quad (3)$$

Notemos que:

$$\sum_{t=0}^{T-1}[f(x^{t+1})-f(x^t)] = \sum_{t=0}^{T-1} f(x^{t+1})-\sum_{t=0}^{T-1} f(x^t) = \sum_{t=1}^{T} f(x^t)-\sum_{t=0}^{T-1} f(x^t) = \sum_{t=1}^{T-1} f(x^t)+f(x^T)-f(x^0)-\sum_{t=1}^{T-1} f(x^t)$$

$$= f(x^T) - f(x^0)$$

Y reemplazando este resultado en (3) tenemos que:

$$f(x^T) - f(x^0) \leq \frac{C_f}{2}\sum_{t=0}^{T-1}\eta_t^2 \quad (4)$$

Ahora como f es convexa y $\mathcal{X}$ es convexo y compacto, entonces $\exists x^* \in \mathcal{X}$ tal que:

$$f(x^*) = min_{x\in\mathcal{X}} f(x)$$

Y como f es diferenciable, entonces $\boxed{\nabla f(x^*) = 0 \quad (5)}$. De este modo:

$$f(x^T) - min_{x\in\mathcal{X}} f(x) = [f(x^T) - f(x^0)] + [f(x^0) - min_{x\in\mathcal{X}} f(x)] = [f(x^T) - f(x^0)] + [f(x^0) - f(x^*)]$$

Y usando (4) llegamos a que:

$$f(x^T) - min_{x\in\mathcal{X}} f(x) = [f(x^T) - f(x^0)] + [f(x^0) - f(x^*)] \leq \frac{C_f}{2}\sum_{t=0}^{T-1}\eta_t^2 + [f(x^0) - f(x^*)]$$

$\Rightarrow$

$$f(x^T) - min_{x\in\mathcal{X}} f(x) \leq \frac{C_f}{2}\sum_{t=0}^{T-1}\eta_t^2 + [f(x^0) - f(x^*)] \quad (6)$$

Utilizando la definición de C_f y usando $\eta_T \in [0,1]$ tenemos que:

$$C_f = \sup_{z,x\in\mathcal{X}:\lambda\in[0,1],y=(1-\lambda)x+\lambda z} \frac{2}{\lambda^2}[f(y) - f(x) - < \nabla f(x), y - x >]$$

$$= \sup_{z,x\in\mathcal{X}:y=(1-\eta_T)x+\eta_T z} \frac{2}{[\eta_T]^2}[f(y) - f(x) - < \nabla f(x), y - x >]$$

Usando $y = x^0$, $x = x^*$ tenemos que:

$$\sup_{z,x\in\mathcal{X}:y=(1-\eta_T)x+\eta_T z} \frac{2}{[\eta_T]^2}[f(y)-f(x)- < \nabla f(x), y-x >] \geq \frac{2}{[\eta_T]^2}[f(x^0)-f(x^*)- < \nabla f(x^*), x^0-x^* >]$$

$\Rightarrow$

$$C_f \geq \frac{2}{[\eta_T]^2}[f(x^0) - f(x^*) - < \nabla f(x^*), x^0 - x^* >]$$

Usando (5) tenemos que:

$$C_f \geq \frac{2}{[\eta_T]^2}[f(x^0) - f(x^*)]$$

$\Rightarrow$

$$\frac{C_f \eta_T^2}{2} \geq f(x^0) - f(x^*)$$

$\Rightarrow$

$$f(x^0) - f(x^*) \leq \frac{C_f \cdot \eta_T^2}{2} \quad (7)$$

Usando (7) en (6) tenemos que:

$$f(x^T) - min_{x\in\mathcal{X}} f(x) \leq \frac{C_f}{2}\sum_{t=0}^{T-1} \eta_t^2 + [f(x^0) - f(x^*)]$$

$$\leq \frac{C_f}{2}\sum_{t=0}^{T-1} \eta_t^2 + \frac{C_f \cdot \eta_T^2}{2} = \frac{C_f}{2}\sum_{t=0}^{T} \eta_t^2$$

$\Rightarrow$

$$f(x^T) - min_{x\in\mathcal{X}} f(x) \leq \frac{C_f}{2}\sum_{t=0}^{T} \eta_t^2 \quad (8)$$

Sea $\eta_{T+1} \in [0, 1]$, entonces:

$$\frac{C_f}{2}\sum_{t=0}^{T} \eta_t^2 = \frac{C_f}{2}\sum_{t=0}^{T} \eta_t^2 + \frac{C_f}{2}\left(\sum_{t=0}^{T} \eta_t^2 + \eta_{T+1}^2\right) = \frac{C_f}{2}\sum_{t=0}^{T+1} \eta_t^2$$

Reemplazando esto en (8) llegamos a que:

$$f(x^T) - min_{x\in\mathcal{X}} f(x) \leq \frac{C_f}{2}\sum_{t=0}^{T} \eta_t^2 \ \leq \frac{C_f}{2}\sum_{t=0}^{T+1} \eta_t^2$$

$\Rightarrow$

$$f(x^T) - min_{x\in\mathcal{X}} f(x) \leq \frac{C_f}{2}\sum_{t=0}^{T+1} \eta_t^2 \quad (9)$$

Si usamos $\eta_t = \frac{2}{2+T} \in [0, 1]$ para todo $t = 0, ..., T + 1$ entonces tenemos que:

$$\frac{C_f}{2}\sum_{t=0}^{T+1} \eta_t^2 = \frac{C_f}{2}\sum_{t=0}^{T+1} \left(\frac{2}{2+T}\right)^2 = \frac{C_f}{2}\sum_{t=0}^{T+1} \frac{4}{(2+T)^2} = \frac{C_f}{2}\cdot\frac{4}{(T+2)^2}\sum_{t=0}^{T+1} 1 = \frac{C_f}{2}\cdot\frac{4}{(T+2)^2}\cdot(T+2) = \frac{2C_f}{T+2}$$

De este modo, si reemplazamos este resultado en (9) llegamos a que:

$$f(x^T) - min_{x\in\mathcal{X}} f(x) \leq \frac{C_f}{2}\sum_{t=0}^{T+1}\eta_t^2 \leq \frac{2C_f}{T+2}$$

Y así finalmente llegamos a que:

$$\boxed{f(x^T) - min_{x\in\mathcal{X}} f(x) \leq \frac{2C_f}{T+2}}$$

Demostrando así lo pedido.

Pregunta 2

a) Esta ejercicio lo resolveremos en 3 partes:

i) Demostraremos C^* es un cono.

ii) Demostraremos que C^* es cerrado.

iii) Demostraremos que $C^* = \{0\} \ \Rightarrow \ C = \mathbf{E}$.

iv) Demostraremos que $C = \mathbf{E} \ \Rightarrow \ C^* = \{0\}$.

---

i) C^* es un cono:

Por definición sabemos que un cono C es un conjunto convexo no vacío tal que $\forall z\lambda \geq 0$ si $z \in C$ entonces $\lambda z \in C$. Y para nuestro caso particular, demostraremos que $C = C^*$ es un cono.

Esta demostración también la divideremos en partes. Lo primero que demostraremos es que C^* es no vacío, luego demostraremos que $\forall \lambda \geq 0$ si $z \in C^*$ entonces $\lambda z \in C^*$ y finalmente demostraremos que C^* es convexo.

Lo primero que debemos demostrar es que C^* es no vacío.

Sabemos que $C^* = \{y \in E :< y, x > \geq 0 \ \forall x \in C\}$.

Considerando que C es no vacío, notemos que $< 0, x >= 0 \ \ \forall x \in C$. De este modo:

$$< 0, x > \geq 0 \ \forall x \in C$$

Así tenemos que $0 \in C^*$ por lo tanto C^* es no vacío.

Ahora demostraremos $\forall \lambda \geq 0$ Si $z \in C^*$ entonces $\lambda z \in C^*$

Sea $z \in C^*$ entonces: $< z, x > \geq 0 \ \forall x \in C$. Sea $\lambda \geq 0$, entonces tenemos que:

$$\lambda < z, x > \geq 0 \ \forall x \in C$$

$\Rightarrow$

$$< \lambda z, x > \geq 0 \ \forall x \in C$$

Sea $u = \lambda z$, entonces:

$$< u, x > \geq 0 \ \forall x \in C$$

$\Rightarrow \quad u \in C^*$. Como $u = \lambda z$, entonces:

Si $z \in C^* \;\Rightarrow \lambda z \in C^*$. Pero como $\lambda \geq 0$ es arbitrario, entonces:

$$\forall \lambda \geq 0 (\text{ Si } z \in C^* \;\Rightarrow\; \lambda z \in C^*)$$

Y como $z \in C^*$ es arbitrario, entonces:

$$\forall z \in C^*, \forall \lambda \geq 0 \;(\text{ Si } z \in C^* \;\Rightarrow\; \forall \lambda z \in C^*)$$

Demostrando así la segunda propiedad que dijimos que demostraríamos.

Sea $\alpha \in [0,1]$ arbitrario entonces tenemos que $\alpha \geq 0$ y que $(1-\alpha) \geq 0$.

Por la segunda propiedad que acabamos de demostrar tenemos que sean $x, y \in C^*$ entonces $\alpha x \in C^*$ y $(1-\alpha)y \in C^*$

Como $\alpha x \in C^*$ y $(1-\alpha)y \in C^*$ entonces:

$$< \alpha x, c_1 > \geq 0 \;\forall c_1 \in C$$

$$< (1-\alpha)y, c_2 > \geq 0 \;\forall c_2 \in C$$

Tomemos un $c \in C$ arbitrario.

Tenemos que:

$$< \alpha x, c > \geq 0 \qquad < (1-\alpha)y, c > \geq 0$$

Si sumamos ambas desigualdades tenemos que:

$$< \alpha x, c > + < (1-\alpha)y, c > \geq 0$$

Como sabemos dado un c fijo, la función $< \cdot, c >$ es lineal. Por ende, la anterior desigualdad implica que:

$$< \alpha x + (1-\alpha)y, c > \geq 0$$

Y como se tomo un $c \in C$ arbitrario, entonces:

$$< \alpha x + (1-\alpha)y, c > \geq 0 \;\forall c \in C$$

De este modo $z = \alpha x + (1-\alpha)y \in C^*$

Así demostramos que si $x, y \in C^*$ entonces $z = \alpha x + (1-\alpha)y \in C^*$.

Y como se tomo un $\alpha \in [0,1]$ arbitrario, entonces:

$$\forall \alpha \in [0,1] \text{ (Si } x, y \in C^* \Rightarrow z = \alpha x + (1-\alpha) y \in C^*)$$

De este modo, se demostró que C^* es convexo.

Así finalmente como demostramos que C^* es un conjunto convexo no vacío y además que $\forall \lambda \geq 0$ (Si $z \in C^* \;\Rightarrow\; \lambda z \in C^*$), entonces demostramos que C^* es un cono, es decir, demostramos todo el inciso i).

---

ii) C^* es cerrado: Sea $z \in \overline{C^*}$, entonces $\exists (z_n)_{n\in\mathbb{N}} \subseteq C^*$ tal que $z_n \;\to z$ cuando $n \to \infty$

Como $(z_n)_{n\in\mathbb{N}} \subseteq C^*$ entonces sea un $\forall n \in \mathbb{N}$ arbitrario, entonces:

$$< z_n, x > \geq 0 \;\; \forall x \in C$$

De este modo, sea un $x \in C$ arbitrario se tiene que:

$$< z_n, x > \geq 0$$

Ahora recordemos que para un $x \in C$ fijo se tiene que la función $< \cdot, x >$ es lineal, y por ende continua.

De este modo:

$$< z, x > = < \lim_{n\to\infty} z_n, x > = \lim_{n\to\infty} < z_n, x > \geq \lim_{n\to\infty} 0 = 0$$

$\Rightarrow$

$$< z, x > \geq 0$$

Como se tomo $x \in C$ arbitrario, entonces:

$$< z, x > \geq 0 \; \forall x \in C$$

De este modo, $z \in C^*$.

Así, se demostró que si $z \in \overline{C^*} \;\Rightarrow z \in C^*$.

De este modo:

$$\overline{C^*} \subseteq C^* \text{ (1)}$$

Y como por definición:

$$C^* \subseteq \overline{C^*} \text{ (2)}$$

Uniendo (1) y (2) tenemos que:

$$C^* = \overline{C^*}$$

Y esto quiere decir que C^* es cerrado.

---

iii) $C^* = \{0\} \;\Rightarrow\; C = \mathbf{E}$:

Sea $A \subseteq E$ un cono arbitrario a partir del cual definimos A^*. Como ya se demostró, A^* es un cono, y como los conos son no vacíos, entonces podemos obtener un $a \in A^*$. Y por definición de cono, tenemos que $\forall \lambda \geq 0 \; \lambda a \in A^*$. En particular si tomamos $\lambda = 0$, entonces $0 = 0 \cdot a = \lambda a \in A^*$. De esta forma $0 \in A^*$. Y esto quiere decir que $\{0\} \subseteq A^*$.

Así, como se había elegido $A \subseteq E$ arbitrario, entonces tenemos que:

$$\boxed{\{0\} \subseteq A^* \;\; \forall A \subseteq E \quad (1)}$$

En particular tomaremos $A = E$, de este modo:

$$\boxed{\{0\} \subseteq E^* \quad (2)}$$

Ahora notemos que sean $A^* \subseteq E$,$B^* \subseteq E$ tales que $A^* \subseteq B^*$, entonces:

Sea $a \in A^*$ arbitrario. Tenemos que $< a, x > \geq 0 \; \forall x \in A$. Como $A^* \subseteq B^*$, entonces esto implica que $< a, x > \geq 0 \; \forall x \in B$. Así tenemos que:

$$< a, x > \geq 0 \; \forall x \in A \;\Rightarrow < a, x > \geq 0 \; \forall x \in B$$

Y esto se da solo si $B \subseteq A$.

De este forma:

$$\boxed{A^* \subseteq B^* \;\Rightarrow\; B \subseteq A \quad (3)}$$

Ahora, sean $A \subseteq E$,$B \subseteq E$ tales que $B \subseteq A$, entonces:

Sea $a \in A^*$ arbitrario. Tenemos que $< a, x > \geq 0 \; \forall x \in A$. Como $B \subseteq A$, entonces esto implica que $< a, x > \geq 0 \; \forall x \in B$. Y esto implica que $a \in B^*$. De este modo:

$$\text{Si } a \in A^* \;\Rightarrow\; a \in B^*$$

Lo cual implica que:

$$A^* \subseteq B^*$$

Así demostramos que:

$$\boxed{B \subseteq A \ \Rightarrow \ A^* \subseteq B^* \quad (4)}$$

Uniendo (4) y (5) tenemos que:

$$\boxed{B \subseteq A \ \Leftrightarrow \ A^* \subseteq B^* \quad (5)}$$

Ahora notemos que usando (5) tenemos que sea $C \subseteq E \ \Rightarrow E^* \subseteq C^*$

Así tenemos que:

$$\boxed{E^* \subseteq C^* \quad (6)}$$

Usando $(2), (6)$ y que $C^* = \{0\}$, entonces:

$$\{0\} \subseteq E^* \subseteq C^* = \{0\}$$

$\Rightarrow$

$$\boxed{E^* = C^* = 0 \quad (7)}$$

Ahora usando (5) tenemos que:

$$(A = B) \Leftrightarrow [(A \subseteq B) \wedge (B \subseteq A)] \Leftrightarrow [(B^* \subseteq A^*) \wedge (A^* \subseteq B^*)] \Leftrightarrow (A^* = B^*)$$

De este modo tenemos que:

$$\boxed{A = B \Leftrightarrow A^* = B^* \quad (8)}$$

Ahora si aplicamos (8) en (7), entonces:

$$C = E$$

De esta manera, se demostró que:

$$\boxed{\text{ Si } C^* = \{0\} \Rightarrow C = E}$$

---

iv) $C = \mathbf{E} \ \Rightarrow \ C^* = \{0\}$:

Sea $C = E$, entonces tenemos que:

$$C^* = \{y \in E :< y, x > \geq 0 \ \forall x \in C = E\} = \{y \in E :< y, x > \geq 0 \ \forall x \in E\}$$

Ahora demostraremos esto por contradicción.

Supongamos que $C^* \neq \{0\}$.

Como sabemos que C^* es no vacío por ser cono, entonces $\exists z \in E : z \neq 0 \wedge z \in C^*$.

Como $z \in C^*$ entonces $< z, x > \geq 0 \forall x \in E$. En particular, si tomamos $x = -z \in E$, entonces:

$$< z, -z > \geq 0 \Rightarrow -\|z\|_2^2 \geq 0 \Rightarrow \|z\|_2^2 \leq 0$$

Y como por definición tenemos que $\|z\|_2^2 \geq 0$, entonces $\|z\|_2^2 = 0 \Leftrightarrow \|z\|_2 = 0 \Leftrightarrow z = 0$, y esto es una contradicción, pues habíamos dicho que $z \neq 0 \rightarrow\leftarrow$.

De este modo, por contradicción se demostró que:

$$\boxed{\text{Si } C = E \Rightarrow C^* = \{0\}}$$

Como se demostró $(i), (ii), (iii)$ y (iv) entonces se demostró que C^* es un cono cerrado y que $C^* = \{0\} \Leftrightarrow C = E$, por lo tanto se demostró a).

---

b) Para demostrar que $(C_1 + C_2)^* = C_1^* \cap C_2^*$, lo que haremos será demostrar que:

i) $(C_1 + C_2)^* \subseteq C_1^* \cap C_2^*$

ii) $C_1^* \cap C_2^* \subseteq (C_1 + C_2)^*$

Lo primero que demostraremos y que se utilizara posteriormente es que $0 \in C$ para todo cono C.

Sea C un cono cualquiera. Como por ser cono es no vacío, entonces $\exists z \in C$. Y como $\forall \lambda \geq 0\ \lambda z \in C$ por el hecho de que C es un cono, entonces tomamos en particular $\lambda = 0$. De este modo:

$$0 = 0z = \lambda z \in C$$

Por lo tanto, para todo cono $C \subseteq E$ tenemos que $0 \in C$.

<u>i) $(C_1 + C_2)^* \subseteq C_1^* \cap C_2^*$</u>: Sea $z \in (C_1 + C_2)^*$, entonces:

$$< z, y > \geq 0\ \forall y \in (C_1 + C_2)$$

Como $y \in (C_1 + C_2)$ entonces se puede escribir $y = x_1 + x_2$ con $x_1 \in C_1$ y $x_2 \in C_2$

De esta manera tenemos que:

$$< z, x_1 + x_2 > \geq 0\ \ \forall x_1 \in C_1, \forall x_2 \in C_2$$

$\Rightarrow$

$$\boxed{< z, x_1 > + < z, x_2 > \geq 0\ \ \forall x_1 \in C_1, \forall x_2 \in C_2 \quad (1)}$$

Recordemos que el 0 pertenece a todo cono. De este modo, utilizamos (1) para reemplazar en el caso particular en que $x_1 = 0$. De este modo, tenemos que:

$$< z, 0 > + < z, x_2 > \geq 0 \quad \forall x_2 \in C_2$$

$\Rightarrow$

$$< z, x_2 > \geq 0 \quad \forall x_2 \in C_2$$

De este modo, tenemos que:

$$\boxed{z \in C_2^* \quad (2)}$$

Nuevamente usaremos que el 0 pertenece a todo cono. De este modo, utilizamos (1) para reemplazar en el caso particular en que $x_2 = 0$. De este modo, tenemos que:

$$< z, x_1 > + < z, 0 > \geq 0 \quad \forall x_1 \in C_1$$

$\Rightarrow$

$$< z, x_1 > \geq 0 \quad \forall x_1 \in C_1$$

De este modo, tenemos que:

$$\boxed{z \in C_1^* \quad (3)}$$

De esta manera, se demostró que:

$$\text{Si } z \in (C_1 + C_2)^* \ \Rightarrow (z \in C_1^*) \land (z \in C_2*)$$

Así tenemos que:

$$\text{Si } z \in (C_1 + C_2)^* \ \Rightarrow z \in C_1^* \cap C_2^*$$

De esta forma:

$$\boxed{(C_1 + C_2)^* \subseteq C_1^* \cap C_2^* \quad (4)}$$

<u>ii) $C_1^* \cap C_2^* \subseteq (C_1 + C_2)^*$</u>: Sea $z \in C_1^* \cap C_2^*$, entonces tenemos que:

$$\boxed{z \in C_1^* \land z \in C_2^* \quad (5)}$$

Como $z \in C_1^*$ entonces:

$$< z, x_1 > \geq 0 \ \forall x_1 \in C_1 \quad (6)$$

Como $z \in C_2^*$ entonces:

$$< z, x_2 > \geq 0 \ \forall x_2 \in C_2 \quad (7)$$

Tomemos un $x \in (C_1 + C_2)$ arbitrario, entonces podemos escribir x como sigue:

$$x = c_1 + c_2 \text{ con } c_1 \in C_1, c_2 \in C_2$$

Como $c_1 \in C_1$, entonces por (6) tenemos que:

$$< z, c_1 > \geq 0 \quad (8)$$

Como $c_2 \in C_2$, entonces por (7) tenemos que:

$$< z, c_2 > \geq 0 \quad (9)$$

Sumando (8) y (9) tenemos que:

$$< z, c_1 > + < z, c_2 > \geq 0$$

Como dado un z fijo $< z, \cdot >$ es lineal, entonces:

$$< z, c_1 + c_2 > \geq 0$$

Y como $x = c_1 + c_2$, entonces:

$$< z, x > \geq 0$$

Como se tomo un $x \in (C_1 + C_2)$ arbitrario, entonces:

$$< z, x > \geq 0 \ \forall x \in (C_1 + C_2)$$

De este modo $z \in (C_1 + C_2)^*$.

Así se demostró que si $z \in C_1^* \cap C_2^* \ \Rightarrow \ z \in (C_1 + C_2)^*$, entonces tenemos que:

$$\boxed{C_1^* \cap C_2^* \subseteq (C_1 + C_2)^* \quad (10)}$$

Como por (4) y (10) tenemos que:

$$[(C_1 + C_2)^* \subseteq C_1^* \cap C_2^*] \wedge [C_1^* \cap C_2^* \subseteq (C_1 + C_2)^*]$$

Entonces tenemos que:

$$\boxed{C_1^* \cap C_2^* = (C_1 + C_2)^*}$$

Que es justamente lo que estábamos buscando demostrar.

---

c) Lo que nos piden demostrar es que $int(C) \neq \emptyset \Leftrightarrow C^*$ es puntiagudo.

Partiremos demostrando $\Rightarrow$.

$\underline{\Rightarrow}$: Aquí partimos de que $int(C) \neq \emptyset$.

Demostraremos esto por contradicción.

Supondremos que C^* no es puntiagudo. Esto significa que $\exists z \neq 0 : z \in C^* \wedge -z \in C^*$

Sea un $c \in C$ arbitrario. Como $z \in C^* \;\Rightarrow\; \boxed{< z, c > \geq 0 \quad (1)}$

Y como $-z \in C^* \;\Rightarrow < -z, c > \geq 0 \;\Rightarrow\; - < z, c > \geq 0 \;\Rightarrow\; \boxed{< z, c > \leq 0 \quad (2)}$

Uniendo (1) y (2) tenemos que $< z, c >= 0$ y como se tomo un $c \in C$ arbitrario entonces:

$$\boxed{< z, c >= 0 \; \forall c \in C \quad (3)}$$

De este modo, podemos decir que z soporta a C en $0 \in C$ pues como $< z, 0 > = 0$, entonces:

$$\boxed{< z, 0 > \leq < z, c > \;\; \forall c \in C}$$

Ahora definimos al hiperplano soportante:

$$H = \{y \in E : < z, y > = 0\}$$

Por (3) es claro que todo punto en C se encuentra en H de este modo:

$$\emptyset \subseteq C \subseteq H$$

$\Rightarrow$

$$int(\emptyset) \subseteq int(C) \subseteq int(H)$$

Usando que $int(\emptyset) = \emptyset$ y que como H es un hiperplano entonces $int(H) = \emptyset$, entonces tenemos que:

$$\emptyset \subseteq int(C) \subseteq \emptyset$$

Por lo tanto:

$$\boxed{int(C) = \emptyset}$$

Lo cual claramente es una contradicción pues habíamos partido suponiendo que $int(C) \neq \emptyset \rightarrow\leftarrow$.

Así demostramos por contradicción que:

$$\boxed{int(C) \neq \emptyset \Rightarrow C^* \text{ es puntiagudo} \quad (4)}$$

$\underline{\Leftarrow}$: Aquí partimos de que C^* es puntiagudo.

Analizaremos casos los posibles casos para que C^* sea puntiagudo.

$\underline{\text{i) } C^* = \{0\}}$ En este caso es claro que C^* es puntiagudo pues $-C^* = \{0\}$ y por ende:

$$[C^*] \cap [-C^*] = \{0\} \cap \{0\} = \{0\}$$

Y como demostramos en a) si $C^* = \{0\} \;\Rightarrow\; C = E$

Esto implica que:

$$\Rightarrow \; int(C) = int(E) \neq \emptyset$$

De este modo, para el caso $C^* = \{0\}$ se demostró que:

$$\boxed{C^* \text{ puntiagudo } \Rightarrow int(C) \neq \emptyset, \text{ para } C^* = \{0\} \quad (5)}$$

ii) $C^* \neq \{0\}$: Como demostramos en a) C^* es siempre no vacío, por ende $\exists z \in C^*$. Y además como también se demostró en a) que C^* es un cono, entonces $\forall \lambda \geq 0 \;\; \lambda z \in C^*$. En particular tomamos $\lambda = 0$, de este modo $0 = 0 \cdot z = \lambda z \in C^*$. Así tenemos que $0 \in C^*$. Por lo tanto:

$$\{0\} \subseteq C^*$$

Pero como estamos tomando $C^* \neq \{0\}$, entonces:

$$\exists z \in C^* : z \neq 0$$

Ahora como C^* es puntiagudo, entonces $[C^*] \cap [-C^*] = \{0\}$ lo que implica que como $z \neq 0$ y $z \in C^*$ entonces $-z \notin C^*$.

Como $-z \notin C^*$ entonces no se cumple que $< -z, c > \geq 0 \; \forall \; c \in C$.

Notemos que $< -z, c > \geq 0 \; \forall \; c \in C$ es equivalente con $< z, c > \leq 0 \; \forall \; c \in C$.

Como esto no se debe cumplir entonces:

$$\boxed{\exists \; c \in C : < z, c > > 0}$$

Como se había tomado un $z \in C^*$ arbitrario, entonces:

$$\boxed{(\exists \; c \in C : < z, c > > 0) \; \forall z \in C^* : z \neq 0 \quad (6)}$$

Ahora, por definición del conjunto C^* tenemos que si $c_1 \in \partial C \;\; \Rightarrow \;\; \exists y \in C^* : c_1 \in H(y)$

Donde $H(y)$ es el hiperplano definido como:

$$H(y) = \{x \in E :< x, y >= 0\}$$

Pero como tenemos por (6) que $c \notin H(z) \; \forall z \in C^*$, entonces $c \notin \partial C$. Pero como $c \in C$, entonces la única opción es que $c \in int(C)$. De este modo, se demuestra que:

$$\boxed{int(C) \neq \emptyset}$$

$\Rightarrow$

$$\boxed{C^* \text{ puntiagudo } \Rightarrow int(C) \neq \emptyset, \text{ para } C^* \neq \{0\} \quad (7)}$$

Uniendo (7) y (5) tenemos que:

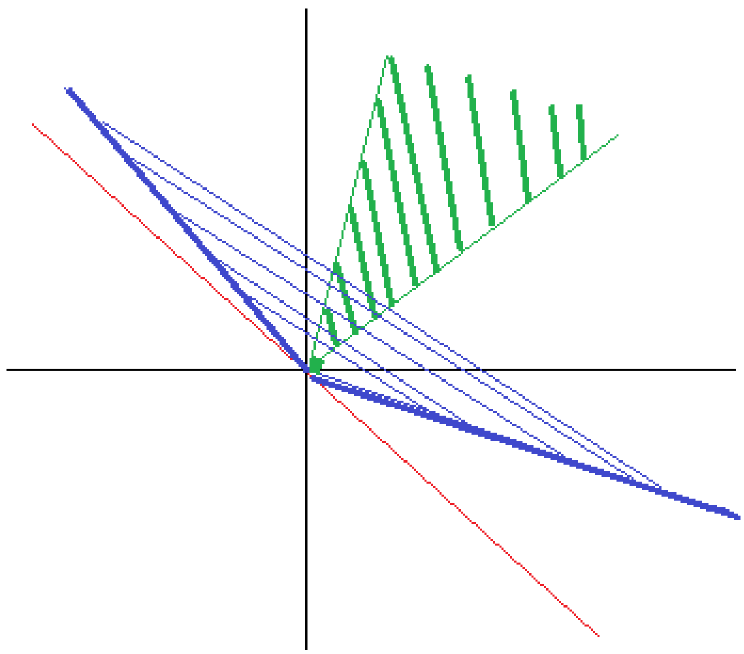
$$\boxed{C^* \text{ puntiagudo } \Rightarrow int(C) \neq \emptyset \quad (8)}$$

Uniendo (8) y (4) tenemos que:

$$\boxed{C^* \text{ puntiagudo } \Leftrightarrow int(C) \neq \emptyset}$$

Que es justamente lo que nos estaban pidiendo demostrar.

Para entender mejor esta situación podría ser de ayuda observar un dibujo en $2D$ de esta situación, en donde el cono azul es C^* y el cono verde es C.



---

d) Lo que nos están pidiendo demostrar es que si C es un cono cerrado, entonces $(C^*)^* = C$.

Lo primero que demostraremos es que $C \subseteq (C^*)^*$.

Sea un $c \in C$ arbitrario.

Tomemos un $c_1 \in C^*$ arbitrario. Tenemos que $< c_1, x > \geq 0 \ \forall x \in C$. De este modo, para el caso particular de $x = c$ tenemos que:

$$< c_1, c > \geq 0$$

$\Rightarrow$

$$< c, c_1 > \geq 0$$

Como tomamos un $c_1 \in C^*$ arbitrario, entonces:

$$< c, c_1 > \geq 0 \forall c_1 \in C^*$$

De este modo $c \in (C^*)^*$. Así demostramos que:

$$\text{Si } c \in C \;\Rightarrow\; c \in (C^*)^*$$

Que es lo mismo a demostrar que:

$$\boxed{C \subseteq (C^*)^* \quad (1)}$$

Esto implica que:

$$\boxed{\overline{C} \subseteq \overline{(C^*)^*}}$$

Pero como por a) tenemos que A^* es cerrado para todo cono $A \subseteq E$, entonces $\overline{(C^*)^*} = (C^*)^*$.

De este modo:

$$\boxed{\overline{C} \subseteq (C^*)^* \quad (2)}$$

Ahora por otro lado tenemos que como $\overline{C}$ es cerrado respecto a E y además como $\overline{C} \subseteq (C^*)^*$ entonces $\overline{C}$ es cerrado respecto a $(C^*)^*$.

Ahora, el complemento de $\overline{C}$ respecto a $(C^*)^*$ entonces:

$$\overline{C}^c = (C^*)^* - (\overline{C})$$

De este modo, como $\overline{C}$ es abierto y convexo, entonces $\overline{C}^c$ es abierto.

Como sabemos por análisis real, la única forma de que si un conjunto A sea convexo, entonces A^c también sea convexo, es que $A = E$ o $A = \emptyset$ o que ∂A sea un hiperplano. Como $\overline{C}$ es un cono, entonces $\partial\overline{C}$ es un hiperplano, de este modo usando esto y que $\overline{C}$ es convexo, entonces $\overline{C}^c$ es convexo.

Definimos $U = \overline{C}^c$ convexo abierto y $V = \overline{C}$ convexo cerrado. Es claro que $\overline{C}^c \cap \overline{C} = \emptyset$. De este modo, podemos utilizar el teorema de separación.

Así tenemos que:

$$\exists w \in (C^*)^* : \;(< u, w >) \;<\; t \;\leq\; (< v, w >) \;\; \forall u \in \overline{C}^c, \forall v \in \overline{C}$$

Notemos que como ya demostramos antes $0 \in \overline{C}$ por ende reemplazando en la desigualdad anterior tenemos que:

$$(< u, w >) \;<\; t \;\leq\; (< 0, w >) \;\; \forall u \in \overline{C}^c$$

$\Rightarrow$

$$(< u, w >) \ < \ t \ \leq \ 0 \ \ \forall u \in \overline{C}^c$$

$\Rightarrow$

$$(< u, w >) \ < 0 \ \ \forall u \in \overline{C}^c$$

De esta forma llegamos a que:

Sean $U = \overline{C}^c$ convexo abierto y $V = \overline{C}$ convexo cerrado (tenemos gratis que son disjuntos), entonces:

$$\boxed{\exists w \in (C^*)^* : \ (< u, w >) \ < \ 0 \ \leq \ (< v, w >) \ \ \forall u \in \overline{C}^c, \forall v \in \overline{C} \quad (3)}$$

Ahora lo que queremos demostrar es que $(C^*)^* \subseteq \overline{C}$.

Lo haremos por contradicción.

Supongamos que no se cumple que $(C^*)^* \subseteq \overline{C}$, de este modo:

$$\boxed{\exists z \in (C^*)^* : z \notin \overline{C} \quad (4)}$$

Por (3) sabemos que $\exists w \in (C^*)^* : w \neq 0$ y que cumple que:

$$(< u, w >) \ < \ 0 \ \leq \ (< v, w >) \ \ \forall u \in \overline{C}^c, \forall v \in \overline{C}$$

Y reemplazando con $u = z$, tenemos que:

$$\boxed{(< z, w >) \ < \ 0 \ \leq \ (< v, w >) \ \forall v \in \overline{C} \quad (5)}$$

Como $0 \ \leq \ (< v, w >) \ \forall v \in \overline{C}$ entonces $w \in (\overline{C})^*$.

$$\boxed{w \in \overline{C}^* \quad (6)}$$

Ahora lo que queremos demostrar es que: $(\overline{C})^* = C^*$

Lo primero que debemos considerar es que: $C \subseteq \overline{C}$ y por las propiedades demostradas en a) tenemos que esto implica que:

$$\boxed{\overline{C}^* \subseteq C^* \quad (7)}$$

Ahora, consideremos un $x \in C^*$ arbitrario.

$$\Rightarrow \ < x, c > \geq 0 \ \forall c \in C$$

Sea un $\overline{c} \in \overline{C}$ arbitrario. Tenemos que $\exists (c_n)_{n \in \mathbb{N}} \subseteq C : c_n \to \overline{c}$ cuando $n \to \ \infty$.

De esta forma:

$$\Rightarrow \ < x, c_n > \geq 0 \ \forall n \in \mathbb{N}$$

Ahora usando esto y que dado un x fijo la función $< x, \cdot >$ es lineal y por ende continua, entonces:

$$< x, \overline{c} > = < x, \lim_{n \to +\infty} c_n > = \lim_{n \to +\infty} < x, c_n > \geq \lim_{n \to +\infty} 0 = 0$$

De esta forma:

$$< x, \overline{c} > \geq 0$$

Y como se tomo un $\overline{c} \in \overline{C}$ arbitrario, entonces:

$$< x, \overline{c} > \geq 0 \ \forall \overline{c} \in \overline{C}$$

$\Rightarrow$

$$x \in \overline{C}^*$$

Así llegamos finamente a que:

$$\text{Si } x \in C^* \Rightarrow x \in \overline{C}^*$$

Y esto implica que:

$$\boxed{C^* \subseteq \overline{C}^* \quad (8)}$$

Así, uniendo (7) y (8) llegamos a que:

$$\boxed{C^* = \overline{C}^* \quad (9)}$$

Así utilizando (9) y (6) llegamos a que:

$$\boxed{w \in C^* \quad (10)}$$

Como por (4) tenemos que $z \in (C^*)^*$ y por (10) tenemos que $w \in C^*$, entonces:

$$\boxed{< z, w > \geq 0 \quad (11)}$$

Pero como por (5) tenemos que $(< z, w >) \ < \ 0$, entonces tenemos una contradicción $\rightarrow\leftarrow$.

De este modo, por contradicción demostramos que:

$$\boxed{(C^*)^* \subseteq \overline{C} \quad (12)}$$

Usando (2) y (12) llegamos a que:

$$\overline{C} \subseteq (C^*)^* \subseteq \overline{C} \quad (13)$$

Lo cual equivale a decir que:

$$\overline{C} = (C^*)^*$$

Y como C es cerrado entonces $\overline{C} = C$ así tenemos que:

$$\boxed{C = (C^*)^*}$$

Que es justamente la igualdad que queríamos demostrar.

---

e) Las cosas que debemos demostrar son que:

i) S_+^d es cerrado.

ii) S_+^d es un cono.

iii) $S_+^d = (S_+^d)^*$.

---

i) Sea un $\overline{A} \in \overline{S_+^d}$, entonces $\exists (A_n)_{n\in\mathbb{N}} \subseteq S_+^d$ tal que $A_n \to \overline{A}$ cuando $n \to \infty$.

Ahora, dado un $x \in \mathbb{R}^d : x \neq 0$ definiremos la siguiente función:

$$g : \overline{S_+^d} \to \mathbb{R} \text{ definida como } g(B) = x^T Bx$$

Ahora, sea $t \in \mathbb{R}$ y $B, C \in \overline{S_+^d}$. Notemos que:

$$g(B+tC) = x^T(B+tC)x = x^T(Bx+tCx) = x^TBx + x^TtCx = x^TBx + t\cdot x^TCx = g(B) + t\cdot g(C)$$

De esta forma g es lineal y por ende continua.

Así, tenemos que:

$$x^T[\overline{A}]x = x^T\left[\lim_{n\to+\infty} A_n\right]x = g\left(\lim_{n\to+\infty} A_n\right) = \lim_{n\to+\infty} g\left(A_n\right) = \lim_{n\to+\infty} x^T A_n\ x$$

$\Rightarrow$

$$\boxed{x^T[\overline{A}]x = \lim_{n\to+\infty} x^T A_n\ x \quad (1)}$$

Como $(A_n)_{n\in\mathbb{N}} \subseteq S_+^d$, entonces:

$$x^T A_n x\ \geq 0\ \ \forall n \in \mathbb{N}$$

Así usando (1) y lo recién mencionado, tenemos que:

$$x^T[\overline{A}]x = \lim_{n \to +\infty} x^T A_n\ x \geq \lim_{n \to +\infty} 0 = 0$$

$$\Rightarrow\ x^T[\overline{A}]x \geq 0$$

Y como se tomo un $x \in \mathbb{R}^d : x \neq 0$ arbitrario, entonces:

$$\boxed{x^T[\overline{A}]x \geq 0\ \forall x \in \mathbb{R}^d : x \neq 0 \quad (2)}$$

Usando (2) tenemos que $\overline{A} \in S_d^+$.

Así, de esta forma tenemos que:

$$\text{Si } \overline{A} \in \overline{S_+^d}\ \Rightarrow\ \overline{A} \in S_+^d$$

Y esto es equivalente a que:

$$\boxed{\overline{S_+^d} \subseteq S_+^d \quad (3)}$$

Y como por definición tenemos que:

$$\boxed{S_+^d \subseteq \overline{S_+^d} \quad (4)}$$

Así usando (3) y (4) tenemos que:

$$\boxed{S_+^d = \overline{S_+^d} \quad (5)}$$

Así tenemos que S_+^d es cerrado.

ii) Para demostrar que es un cono, lo primero que demostraremos es que es no vacío, pues la matriz en la que todos sus elementos son solo 0's pertenece a S_+^d.

Ahora sea $A \in S_+^d$ tenemos que todos sus valores propios son mayores o iguales a 0. Es decir:

$$Av_i = \lambda_i v_i \ \text{ con }\ \lambda_i \geq 0\ \forall i \in \{1, 2, ..., d\}$$

De esta manera, sea un $t \geq 0$ entonces:

$$[t \cdot A]v_i = t \cdot [Av_i] = [t \cdot \lambda_i]v_i \ \ \forall i \in \{1, 2, ..., d\}$$

Así tenemos que la matriz $t \cdot A$ tiene asociados los valores propios $t\lambda_1, t\lambda_2, ..., t\lambda_d$

Pero como $t \geq 0$ y $\lambda_i \geq 0\ \forall i \in \{1, ..., d\}$, entonces $t \cdot \lambda_i \geq 0\ \forall i \in \{1, ..., d\}$

De esta manera todo los valores propios de tA son positivos.

Además como tenemos que sean $(a_{i,j})_{i,j\in\{1,2,...,d\}}$ los elementos de A entonces como A es simétrica tenemos que:

$$a_{i,j} = a_{j,i} \ \forall i,j \in \{1,...,d\}$$

Como los elementos de tA estarán dados por $(t \cdot a_{i,j})_{i,j\in\{1,2,...,d\}}$ entonces tenemos que:

$$[tA]_{i,j} = ta_{i,j} = ta_{j,i} = [tA]_{j,i} \quad \forall i,j \in \{1,2,...,d\}$$

Y así tenemos que tA es simétrica, y como todos sus valores propios son positivos, entonces:

$$\text{Si } A \in S_+^d \ \Rightarrow \ tA \in S_+^d$$

Como se escogió un $t \geq 0$ arbitrario, entonces:

$$\boxed{\text{Si } A \in S_+^d \ \Rightarrow \ [(tA \in S_+^d) \ \forall t \geq 0] \ (6)}$$

Sean $A, B \in S_+^d$ tenemos por definición que:

$$\forall x \in \mathbb{R}^d : x \neq 0 \quad x^T Ax \geq 0 \text{ y } x^T Bx \geq 0$$

De este modo, sea un $x \in \mathbb{R}^d : x \neq 0$ arbitrario, entonces:

$$x^T Ax \geq 0 \ \text{ y } \ x^T Bx \geq 0$$

Así tenemos que:

$$x^T(A+B)x = x^T(Ax+Bx) = x^T Ax + x^T Bx \geq 0$$

De esta forma:

$$x^T(A+B)x \geq 0$$

Y como se tomo un $x \in \mathbb{R}^d : x \neq 0$ arbitrario, entonces:

$$x^T(A+B)x \geq 0 \ \forall x \in \mathbb{R}^d : x \neq 0$$

Lo que quiere decir que $(A+B)$ es semidefinida positiva.

Ahora como sabemos que tanto A como B son simétricas, entonces:

$$[A+B]_{i,j} = a_{i,j} + b_{i,j} = a_{j,i} + b_{j,i} = [A+B]_{j,i} \ \forall i,j \in \{1,2,..,d\}$$

Así tenemos que $A+B$ es simétrica.

Como tenemos que $A+B$ es simétrica y semidefinida positiva, entonces:

$$\boxed{\text{Si } A, B \in S_+^d \ \Rightarrow \ (A+B) \in S_+^d \quad (7)}$$

Ahora sea un $\alpha \in [0, 1]$ arbitrario. Tenemos que $(1 - \alpha) \geq 0$ y $\alpha \geq 0$.

Sean $A, B \in S^d_+$ entonces por (6) tenemos que:

$$\alpha A \in S^d_+ \text{ y } (1 - \alpha)B \in S^d_+$$

Y por (7) tenemos que:

$$[\alpha A + (1 - \alpha)B] \in S^d_+$$

Y como se tomo un $\alpha \in [0, 1]$ arbitrario, entonces:

$$([\alpha A + (1 - \alpha)B] \in S^d_+)\forall\ \alpha \in [0, 1]$$

Así tenemos que S^d_+ es convexo.

Como se demostró que S^d_+ es convexo y no vacío y usando lo demostrado en (6), es decir, que si $A \in S^d_+ \ \Rightarrow\ \forall\ t \geq 0(\ t \cdot A \in S^d_+)$, entonces demostramos que S^d_+ es un cono.

iii) Sea $A \in S^d_+$ consideremos su descomposición SVD:

$$A = USV^T$$

con U, V matrices ortonormales y S matriz diagonal que tiene sus valores singulares ordenados de mayor a menor.

Como sabemos, los valores singulares son la raiz de los valores propios de A^TA. Pero como A es simetrica entonces $A^TA = AA = A^2$. Así tenemos que los valores singulares de la matriz A son la raiz de sus valores propios al cuadrado. Es decir $\sigma_i = \sqrt{[\lambda_i]^2} = |\lambda_i|\ \forall i \in \{1, ..., d\}$.

Pero como A es semidefinida positiva, entonces $\lambda_i \geq 0\ \forall i \in \{1, ..., d\} \ \Rightarrow\ \sigma_i = |\lambda_i| = \lambda_i\ \forall i \in \{1, ..., d\}$.

De este modo:

$$S = diag(\lambda_1, \lambda_2, ..., \lambda_d)$$

Ahora usando nuevamente que A es simétrica tenemos que:

$$USV^T = A = A^T = (USV^T)^T = (V^T)^TS^TU^T = VSU^T$$

$\Rightarrow\ USV^T = VSU^T\ \Rightarrow U = V$. Así finalmente tenemos que:

$$A = VSV^T$$

Con $S = diag(\lambda_1, \lambda_2, ..., \lambda_d)$ y V matriz ortonormal.

De esta forma, si $A \in S^d_+$ entonces se puede escribir como sigue:

$$\boxed{A = \sum_{i=1}^{d} \lambda_i^A v_i^A (v_i^A)^T \quad (8)}$$

Ahora consideremos una matriz $A \in \mathbb{R}^{d\times d}$ arbitraria y su descomposición SVD:

$$A = U^A S^A (V^A)^T$$

Así tenemos que:

$$A = \sum_{i=1}^{d} (u_i^A)\sigma_i^A (v_i^A)^T$$

Y consideremos un $B \in S_+^d$ arbitrario, entonces por (8) tenemos que:

$$B = \sum_{j=1}^{d} \lambda_j^B v_j^B (v_j^B)^T$$

$$A^T B = \left(\sum_{i=1}^{d} (u_i^A)\sigma_i^A (v_i^A)^T\right)^T \left(\sum_{j=1}^{d} \lambda_j^B v_j^B (v_j^B)^T\right) = \left(\sum_{i=1}^{d} [(v_i^A)^T]^T \sigma_i^A (u_i^A)^T\right) \left(\sum_{j=1}^{d} \lambda_j^B v_j^B (v_j^B)^T\right)$$

$$= \left(\sum_{i=1}^{d} (v_i^A)\sigma_i^A (u_i^A)^T\right) \left(\sum_{j=1}^{d} \lambda_j^B v_j^B (v_j^B)^T\right) = \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_i^A \lambda_j^B (v_i^A)(u_i^A)^T v_j^B (v_j^B)^T$$

$$= \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_i^A \lambda_j^B (v_i^A)[(u_i^A)^T v_j^B](v_j^B)^T = \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_i^A \lambda_j^B [(u_i^A)^T v_j^B](v_i^A)(v_j^B)^T$$

$\Rightarrow$

$$\boxed{A^T B = \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_i^A \lambda_j^B [(u_i^A)^T v_j^B](v_i^A)(v_j^B)^T \quad (9)}$$

Ahora notemos que:

$$(u_i^A)^T v_j^B = \sum_{t=1}^{d} [u_i^A]_t [v_j^B]_t$$

Y reemplazando en (9) llegamos a que:

$$\boxed{A^T B = \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_i^A \lambda_j^B \left(\sum_{t=1}^{d} [u_i^A]_t [v_j^B]_t\right)(v_i^A)(v_j^B)^T \quad (10)}$$

De este modo, definimos:

$$x_{i,j} = \sum_{t=1}^{d} [u_i^A]_t [v_j^B]_t$$

Y así tenemos que reemplazando con este valor en (10) llegamos a que:

$$\boxed{A^TB = \sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_i^A\lambda_j^B x_{i,j}(v_i^A)(v_j^B)^T \quad (11)}$$

Ahora notemos que:

$$< A,B > = \ traza(A^TB) = traza\left[\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_i^A\lambda_j^B x_{i,j}(v_i^A)(v_j^B)^T\right] = \sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_i^A\lambda_j^B x_{i,j}{\cdot}traza\left[(v_i^A)(v_j^B)^T\right] \quad (12)$$

Es claro que:

$$[(v_i^A)(v_j^B)^T]_{k,k} = (v_i^A)_k(v_j^B)_k$$

De esta forma:

$$traza\left[(v_i^A)(v_j^B)^T\right] = \sum_{k=1}^{d}(v_i^A)_k(v_j^B)_k$$

Ahora definimos:

$$y_{i,j} = \sum_{k=1}^{d}(v_i^A)_k(v_j^B)_k$$

Reemplazando en (12) llegamos a que:

$$< A,B > = \sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_i^A\lambda_j^B x_{i,j}\cdot traza\left[(v_i^A)(v_j^B)^T\right] = \sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_i^A\lambda_j^B x_{i,j}\cdot y_{i,j}$$

Finalmente llegamos a que sea un $A \in \mathbb{R}^{d\times d}$ arbitrario, entonces:

$$\boxed{< A,B > = \ \sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_i^A\lambda_j^B x_{i,j}\cdot y_{i,j} \quad (13)}$$

Con $\sum_{t=1}^{d}[u_i^A]_t[v_i^B]_t$ y $y_{i,j} = \sum_{k=1}^{d}(v_i^A)_k(v_i^B)_k$

Ahora comenzaremos con la demostración propiamente tal.

Sea un $A \in (S_d)^*$ arbitrario, entonces:

$$< A,B > \geq 0 \ \ \forall B \in S_+^d$$

Usando (13) tenemos que para que esta desigualdad se cumpla para todo $B \in S_+^d$ entonces debe ocurrir que $x_{i,j} = y_{i,j}$ $\forall i,j \in \{1,...,d\}$.

Y para que $\forall B \in S_+^d$ se de que $x_{i,j} = y_{i,j}$ es necesario que $u_i^A = v_i^A \ \ \forall i \in \{1,...,d\}$.

De este modo, tenemos que $V^A = U^A$

Y como la factorización para A era:

$$A = U^A S^A (V^A)^T$$

Entonces tenemos que:

$$A = V^A S^A (V^A)^T$$

Así tenemos que la matriz A es simétrica.

Además de eso, esta factorización de A nos dice que sus valores propios son iguales a sus valores singulares, y como todos los valores singulares son mayor o igual a 0, entonces todos los valores propios de A son mayor o igual a 0, entonces A es semidefinida positiva.

Así como A es simétrica y semidefinida positiva, entonces:

$$A \in S_+^d$$

De este modo, se mostró que:

$$\text{Si } A \in (S_+^d)^* \;\Rightarrow\; A \in S_+^d$$

Lo cual es equivalente con que:

$$\boxed{(S_+^d)^* \subseteq S_+^d \quad (14)}$$

Sea un $A \in S_+^d$ arbitrario.

Si consideremos un $B \in S_+^d$, tenemos por (13) que:

$$< A, B > = \sum_{i=1}^{d} \sum_{j=1}^{d} \sigma_i^A \lambda_j^B x_{i,j} \cdot y_{i,j}$$

con $\sum_{t=1}^{d} [u_i^A]_t [v_i^B]_t$ y $y_{i,j} = \sum_{k=1}^{d} (v_i^A)_k (v_i^B)_k$

Como $A \in S_+^d$, entonces $u_i^A = v_i^A \ \forall i \in \{1, ..., d\}$ lo cual implica que $x_{i,j} = y_{i,j} \ \forall i, j \in \{1, ..., d\}$

Además como $A \in S_+^d$ por lo demostrado en (8) tenemos que $\sigma_i^A = \lambda_i^A \ \forall i \in \{1, ..., d\}$

De este modo:

$$< A, B > = \sum_{i=1}^{d} \sum_{j=1}^{d} \sigma_i^A \lambda_j^B x_{i,j} \cdot y_{i,j} = \sum_{i=1}^{d} \sum_{j=1}^{d} \lambda_i^A \lambda_j^B x_{i,j} \cdot x_{i,j} = \sum_{i=1}^{d} \sum_{j=1}^{d} \lambda_i^A \lambda_j^B (x_{i,j})^2$$

Así como tanto $\lambda_i^A \geq 0 \forall i \in \{1, ..., d\}$ y $\lambda_j^B \geq 0 \forall j \in \{1, ..., d\}$ entonces:

$$< A, B >= \sum_{i=1}^{d} \sum_{j=1}^{d} \lambda_i^A \lambda_j^B (x_{i,j})^2 \geq 0$$

$\Rightarrow$

$$< A, B > \geq 0$$

Y como se tomo un $B \in S_+^d$ arbitrario, entonces:

$$< A, B > \geq 0 \ \ \forall B \in S_+^d$$

$$\Rightarrow A \in (S_+^d)^*$$

Así como partimos de que $A \in S_+^d$ arbitrario, entonces demostramos que:

$$\text{Si } A \in S_+^d \Rightarrow A \in (S_+^d)^*$$

Lo cual implica que:

$$\boxed{S_+^d \subseteq (S_+^d)^* \quad (15)}$$

Uniendo (14) y (15) llegamos a que:

$$(S_+^d)^* \subseteq S_+^d \subseteq (S_+^d)^*$$

Lo cual implica que:

$$\boxed{(S_+^d)^* = S_+^d}$$

Así, como demostramos (i), (ii) y (iii) entonces demostramos que:

$$\boxed{S_+^d \text{ es un cono cerrado y que } (S_+^d)^* = S_+^d}$$

Lo cual es todo lo correspondiente al inciso e).

---

f) Lo primero que debemos hacer es demostrar que sea $y \in C^*$, entonces el conjunto:

$K = \{x \in C :< y, x >= 1\}$ es compacto.

Sea $\overline{k} \in \overline{K} \ \Rightarrow \ \exists (k_n)_{n \in \mathbb{N}} \subseteq K : k_n \to \overline{k}$ cuando $n \to \infty$

Como $(k_n)_{n \in \mathbb{N}} \subseteq K \ \Rightarrow < y, k_n >= 1 \ \forall n \ \in \mathbb{N} \ \Rightarrow \ \text{lím}_{n \to \infty} < y, k_n >= \text{lím}_{n \to \infty} 1 = 1$

$\Rightarrow \ \text{lím}_{n \to \infty} < y, k_n >= 1$

Ahora, notemos que dado un $y \in C^*$ la función $< y, \cdot >$ es lineal y por ende continua. De este modo:

$\Rightarrow < y, \text{lím}_{n \to \infty} k_n >= 1$

$\Rightarrow < y, \overline{k} >= 1 \ \Rightarrow \ \overline{k} \in K$

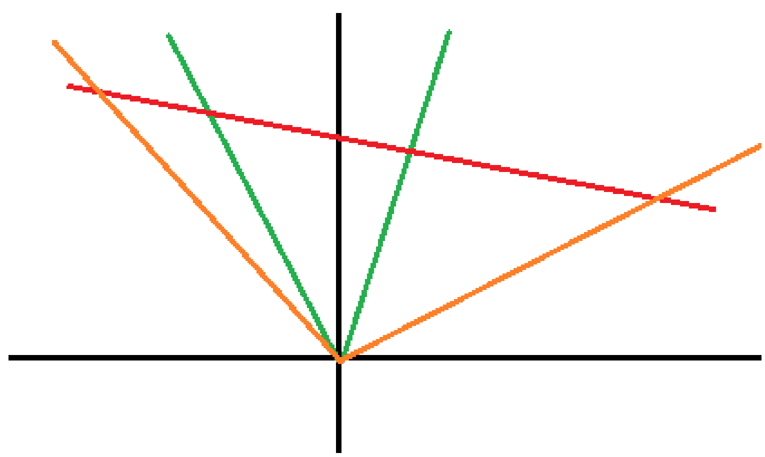
De este modo, demostramos que si $\overline{k} \in \overline{K} \;\Rightarrow\; \overline{k} \in K$

Que es equivalente a decir que:

$\overline{K} \subseteq K$ y como por definición $K \subseteq \overline{K}$, entonces $\boxed{K = \overline{K}}$

Así tenemos que K es cerrado.

Por otro lado, tenemos que el conjunto K es como un corte que se le hace al cono (como se ve en la figura). El cono naranjo es C^*, mientras que el cono verde es C y la linea roja es K.



Ahora como tenemos que como C es puntiagudo, entonces K es acotado.

Como se tiene que K es cerrado y acotado, entonces K es compacto.

Ahora, sean $k_1, k_2 \in K$, entonces tenemos que $< y, k_1 > = 1$ y $< y, k_2 > = 1$

Ahora tomemos un $\alpha \in [0,1]$, entonces definimos $k_3 = \alpha k_1 + (1-\alpha)k_2$.

De este modo, tenemos que:

$< y, k_3 > = < y, \alpha k_1 + (1-\alpha)k_2 > = < y, \alpha k_1 > + < y, (1-\alpha)k_2 >$

$= \; \alpha < y, k_1 > + (1-\alpha) < y, k_2 > = \; \alpha + (1-\alpha) = 1$

De este modo:

$< y, k_3 > = 1 \;\Rightarrow\; k_3 \in K$

De este modo, se demuestra que K es convexo.

Como K es convexo y compacto, entonces:

$K = conv(ext(K))$

Ahora, sea $x \in ext(K)$, entonces:

$$\text{Si } x = \frac{a+b}{2} \text{ con } a, b \in K \ \Rightarrow \ a = b = x$$

$\Leftrightarrow$

$$\text{Si } 2x = a+b \text{ con } a, b \in K \ \Rightarrow \ a = b = x$$

$\Leftrightarrow$

Sea $u = 2x$, entonces:

$$\text{Si } u = a+b \text{ con } a, b \in K \ \Rightarrow \ a = b = \frac{u}{2} \ \Rightarrow \ a = \frac{u}{2} \ \Rightarrow \ \exists \alpha = \frac{1}{2} \geq 0 : a = \alpha u$$

De esta forma $\{\alpha u : \alpha \geq 0\}$ es un rayo, que es lo mismo a decir que $\{\alpha 2x : \alpha \geq 0\}$ es un rayo.

Si tomamos $\alpha_1 = 2\alpha$, entonces $\{\alpha \cdot 2x : \alpha \geq 0\}$ es el mismo conjunto que:

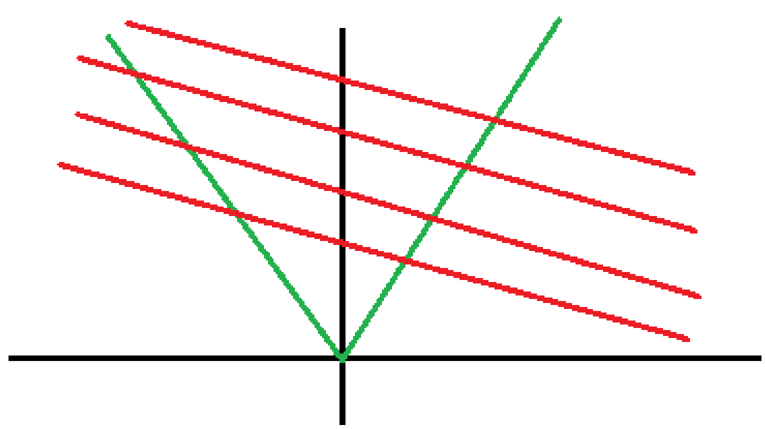
$\{\alpha_1 x : \alpha_1 \geq 0\}$

De esta forma $\{\alpha_1 x : \alpha_1 \geq 0\}$ es un rayo extremo.

Así, demostramos que sea $x \in ext(K)$ entonces $\{\alpha_1 x : \alpha_1 \geq 0\}$ es un rayo extremo.

Ahora definimos a los conjuntos:

$$K_t = \{x \in C : < y, x > = t\} \ \ \forall \ \ t \geq 0$$



Se graficaron los conjuntos k_t en la imagen recién presentada.

También notemos que sean $t_1 \geq 0, t_2 \geq 0 : t_1 \neq t_2$ entonces $k_{t_1} \cap k_{t_2} = \emptyset$

Ahora tenemos que:

Sea $x \in C \Leftrightarrow \exists t \geq 0 : x \in K_t \;\Leftrightarrow\; \exists l_1, ..., l_n \in ext(K_t) : x = \sum_{i=1}^{n} \alpha_i \cdot l_i \;\Leftrightarrow\; \exists l_1, ..., l_n \in rext(C) : x = \sum_{i=1}^{n} \alpha_i \cdot l_i \;\Leftrightarrow\; x \in conv(rext(C))$

De esta forma $x \in C \Leftrightarrow x \in conv(rext(C))$, así tenemos que:

$$\boxed{C = conv(rext(C))}$$

---

g) Consideremos una matriz $A \in S_+^d$ como ya demostramos antes, esto implica que A se puede escribir de la siguiente forma:

$$A = \sum_{i=1}^{d} \lambda_i v_i^A (v_i^A)^T$$

Ahora notemos que sea un $x \in \mathbb{R}^d$ arbitrario, entonces para un $i \in \{1, ..., d\}$ podemos escribir:

$$v_i = (v_i - x) + x$$

Definimos $y_i = v_i - x$, de esta forma:

$$v_i = y_i + x$$

$$(v_i)(v_i)^T = [y_i + x][y_i + x]^T = [y_i + x][(y_i)^T + (x)^T] = (y_i)(y_i)^T + (y_i)x^T + x(y_i)^T + xx^T$$

Multiplicamos a ambos lados por λ_i y llegamos a que:

$$\lambda_i(v_i)(v_i)^T = \lambda_i(y_i)(y_i)^T + \lambda_i(y_i)x^T + \lambda_i x(y_i)^T + \lambda_i xx^T$$

$\Rightarrow$

$$\begin{aligned} A = \sum_{i=1}^{d} \lambda_i(v_i)(v_i)^T &= \sum_{i=1}^{d} \left[\lambda_i(y_i)(y_i)^T + \lambda_i(y_i)x^T + \lambda_i x(y_i)^T + \lambda_i xx^T\right] \\ &= \sum_{i=1}^{d}[\lambda_i(y_i)(y_i)^T] + \sum_{i=1}^{d}[\lambda_i y_i x^T] + \sum_{i=1}^{d} \lambda_i x y_i^T + \left(\sum_{i=1}^{d} \lambda_i\right) xx^T \end{aligned}$$

$\Rightarrow$

$$\boxed{A = \sum_{i=1}^{d}[\lambda_i(y_i)(y_i)^T] + \sum_{i=1}^{d}[\lambda_i y_i x^T] + \sum_{i=1}^{d} \lambda_i x y_i^T + \left(\sum_{i=1}^{d} \lambda_i\right) xx^T \quad (1)}$$

Ahora como vamos a tomar $A, B \in S_+^d$ entonces por (1) tenemos que:

$$\boxed{A = \sum_{i=1}^{d}[\lambda_i^A (y_i^A)(y_i^A)^T] + \sum_{i=1}^{d}[\lambda_i^A y_i^A x^T] + \sum_{i=1}^{d} \lambda_i^A x (y_i^A)^T + \left(\sum_{i=1}^{d} \lambda_i^A\right) xx^T \quad (2)}$$

$$\boxed{B = \sum_{i=1}^{d}[\lambda_i^B (y_i^B)(y_i^B)^T] + \sum_{i=1}^{d}[\lambda_i^B y_i^B x^T] + \sum_{i=1}^{d} \lambda_i^B x (y_i^B)^T + \left(\sum_{i=1}^{d} \lambda_i^B\right) xx^T \quad (3)}$$

Ahora lo que haremos será decir que tenemos que se cumple que:

$$xx^T = A + B \quad (4)$$

Entonces tenemos que:

$$xx^T = \sum_{i=1}^{d}[\lambda_i^A (y_i^A)(y_i^A)^T] + \sum_{i=1}^{d}[\lambda_i^A y_i^A x^T] + \sum_{i=1}^{d} \lambda_i^A x (y_i^A)^T + \left(\sum_{i=1}^{d} \lambda_i^A\right) xx^T +$$
$$\sum_{i=1}^{d}[\lambda_i^B (y_i^B)(y_i^B)^T] + \sum_{i=1}^{d}[\lambda_i^B y_i^B x^T] + \sum_{i=1}^{d} \lambda_i^B x (y_i^B)^T + \left(\sum_{i=1}^{d} \lambda_i^B\right) xx^T$$

Notemos que como tenemos solo matrices de rango (1) entonces lo que debe ocurrir es que:

$$y_i^A = c_i^A x \ \ \forall i \in \{1, ..., d\} \quad (5) \quad y_i^B = c_i^B x \ \ \forall i \in \{1, ..., d\} \quad (6)$$

Y como habíamos definido que:

$$y_i^A = v_i^A - x \ \Rightarrow \ \ [\text{por } (5)] \ \ c_i^A x = v_i^A - x \ \Rightarrow \ v_i^A = (c_i^A + 1)x \quad (7)$$

$$y_i^B = v_i^B - x \ \Rightarrow \ \ [\text{por } (6)] \ \ c_i^B x = v_i^B - x \ \Rightarrow \ v_i^B = (c_i^B + 1)x \quad (8)$$

Ahora volviendo a usar la descomposición de A como suma de d matrices de rango 1 (por factorización SVD) y usando (7) y (8) tenemos que:

$$A = \sum_{i=1}^{d} \lambda_i^A v_i^A (v_i^A)^T = \sum_{i=1}^{d} \lambda_i^A [(c_i^A+1)x][(c_i^A+1)x]^T = \sum_{i=1}^{d} \lambda_i^A (c_i^A+1)^2 xx^T = xx^T \left(\sum_{i=1}^{d} \lambda_i^A (c_i^A + 1)^2\right)$$

Sea $\alpha_A = \sum_{i=1}^{d} \lambda_i^A (c_i^A + 1)^2$, entonces tenemos que:

$$A = \alpha_A (xx^T) \quad (9)$$

$$B = \sum_{i=1}^{d} \lambda_i^B v_i^B (v_i^B)^T = \sum_{i=1}^{d} \lambda_i^B [(c_i^B+1)x][(c_i^B+1)x]^T = \sum_{i=1}^{d} \lambda_i^B (c_i^B+1)^2 xx^T = xx^T \left(\sum_{i=1}^{d} \lambda_i^B (c_i^B + 1)^2\right)$$

Sea $\alpha_B = \sum_{i=1}^{d} \lambda_i^B (c_i^B + 1)^2$, entonces tenemos que:

$$B = \alpha_B(xx^T) \quad (10)$$

Ahora usando lo que habiamos planteado en (4) junto con (9) y (10) entonces tenemos que:

$$xx^T = \alpha_A(xx^T) + \alpha_B(xx^T) \quad (11)$$

Entonces tenemos que:

$$\alpha_A + \alpha_B = 1 \quad (12)$$

Ahora como $A \in S_+^d$ y como $A = \alpha_A \;\Rightarrow\; \alpha_A \geq 0 \quad (13)$

Y como $B \in S_+^d$ y como $B = \alpha_B \;\Rightarrow\; \alpha_B \geq 0 \quad (14)$

Ahora diremos que $\alpha_B = \alpha \;\Rightarrow\;$ [por (12)] $\alpha_A = (1-\alpha) \quad (15)$

Por (14) tenemos que $\alpha \geq 0 \quad (16)$ y por (13) utilizado en (15) tenemos que $1-\alpha \geq 0 \;\Rightarrow\; 1 \geq \alpha \quad (17)$

Combinando (16) y (17) tenemos que $\alpha \in [0,1] \quad (18)$.

Así llegamos a que $A = (1-\alpha)xx^T \quad (19)$ y como por (18) $\alpha \in [0,1] \;\Rightarrow\; (1-\alpha) \geq 0$, entonces tenemos que:

$$\text{Si } x^T x = A + B \;\Rightarrow\; \exists t \geq 0 : A = txx^T \text{ en particular tomamos } t = (1-\alpha)$$

De este modo, usando la definición que se presenta en el enunciado de la pregunta 2, inciso f), demostramos que los rayos de S_+^d son conjuntos de la forma:

$$\{txx^T : t \geq 0\}$$

Que es justamente lo que nos habían pedido.

Pregunta 3

Para esta sección lo primero que haremos será mostrar como llegamos a la expresión de cada algoritmo (Parte teórica de los algoritmos) y luego de eso se presentaran los gráficos y todo el análisis pertinente.

Parte Teórica de los algoritmos

a) Notemos que lo queremos aplicar es el método del gradiente proximal. Y ese método funciona de la siguiente manera:

1. Sea $x^0 \in \mathcal{X}$ dado y T dado.

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

   $x^{t+1} = argmin_{x\in\mathcal{X}}\{f(x^t) + < \nabla f(x^t), x - x^t > + \frac{1}{2\eta_t}\|x - x^t\|^2\}$

3. return x^T

Ahora notemos que como estamos usando norma 2, entonces:

$$x^{t+1} = argmin_{x\in\mathcal{X}}\{f(x^t) + < \nabla f(x^t), x - x^t > + \frac{1}{2\eta_t}\|x - x^t\|_2^2\}$$

$$= argmin_{x\in\mathcal{X}}\{< \nabla f(x^t), x - x^t > + \frac{1}{2\eta_t}\|x - x^t\|_2^2\}$$

$$= argmin_{x\in\mathcal{X}}\{\frac{1}{2\eta_t}\left[2\eta_t < \nabla f(x^t), x - x^t > + \|x - x^t\|_2^2\right]\}$$

$$= argmin_{x\in\mathcal{X}}\{2\eta_t \cdot < \nabla f(x^t), x - x^t > + \|x - x^t\|_2^2\}$$

$$= argmin_{x\in\mathcal{X}}\{\|\eta_t \cdot \nabla f(x^t)\|_2^2 + 2\eta_t \cdot < \nabla f(x^t), x - x^t > + \|x - x^t\|_2^2\}$$

$$= argmin_{x\in\mathcal{X}}\{< \eta_t \cdot \nabla f(x^t), \eta_t \cdot \nabla f(x^t) > + 2 \cdot < \eta_t \cdot \nabla f(x^t), x - x^t > + < x - x^t, x - x^t >\}$$

$$= argmin_{x\in\mathcal{X}}\{\|x - x^t + \eta_t \nabla f(x^t)\|_2^2\} = argmin_{x\in\mathcal{X}}\{\|x - [x^t - \eta_t \nabla f(x^t)]\|_2^2\}$$

$$= \Pi_{\mathcal{X}}(x^t - \eta_t \nabla f(x^t))$$

De esta forma, el algoritmo nos queda como sigue:

1. Sea $x^0 \in \mathcal{X}$ dado y T dado.

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

   $x^{t+1} = \Pi_{\mathcal{X}}(x^t - \eta_t \nabla f(x^t))$

3. return x^T

Ahora tenemos que calcular la proyección a una bola unitaria respecto a la norma 1. Este problema es:

$$min\{\|x - \overline{x}\|_2^2 : \|x\|_1 \leq 1\}$$

Que es lo mismo a que:

$$min\{\sum_{i=1}^{d}[x_i - \overline{x_i}]^2 : \sum_{i=1}^{d}|x_i| \leq 1\}$$

Si hacemos la sustitución $x_i = signo(x_i) \cdot t_i$ con $t_i \geq 0$ $\forall i \in \{1, ..., d\}$ entonces nos queda que:

$|x_i| = |signo(x_i) \cdot t_i| = |signo(x_i)| \cdot |t_i| = t_i$

Así la restricción queda como:

$$\sum_{i=1}^{d} t_i \leq 1$$

Ademas, notemos que la función objetivo queda como sigue:

$$\sum_{i=1}^{d}[x_i - \overline{x_i}]^2 = \sum_{i=1}^{d}[signo(x_i) \cdot t_i - \overline{x_i}]^2 = \sum_{i=1}^{d}[signo(x_i) \cdot t_i - \overline{x_i}]^2$$

$$= \sum_{i=1}^{d}([signo(x_i)]^2 \cdot {t_i}^2 - 2signo(x_i) \cdot t_i \cdot \overline{x_i} + [\overline{x_i}]^2)$$

$$= \sum_{i=1}^{d}({t_i}^2 - 2signo(x_i) \cdot t_i \cdot \overline{x_i} + [\overline{x_i}]^2)$$

Notemos que elegimos $signo(x_i) = signo(\overline{x_i})$ para así minimizar, pues es el valor que hace que la expresión del medio siempre sea negativa. De este modo, nos queda:

$$= \sum_{i=1}^{d}({t_i}^2 - 2 \cdot t_i \cdot |\overline{x_i}| + [|\overline{x_i}|]^2) = \sum_{i=1}^{d}(t_i - |\overline{x_i}|)^2$$

Así el problema es equivalente a:

$$min\{\sum_{i=1}^{d}(t_i - |\overline{x_i}|)^2 : \sum_{i=1}^{d} t_i \leq 1, t \geq 0\}$$

Planteamos el lagrangeano:

$$L(\lambda, t) = \sum_{i=1}^{d}(t_i - |\overline{x_i}|)^2 + \lambda \left(\sum_{i=1}^{d} t_i - 1\right)$$

Ahora calculamos la derivada parcial respecto a t_k y nos queda que:

$$\frac{\partial L(\lambda, t)}{\partial t_k} = 2 \cdot (t_k - |\overline{x_k}|) + \lambda$$

De esta forma en el optimo:

$$2 \cdot (t_k - |\overline{x_k}|) + \lambda = 0 \;\Rightarrow\; 2 \cdot (t_k - |\overline{x_k}|) = -\lambda \;\Rightarrow\; t_k - |\overline{x_k}| = -\frac{\lambda}{2} \;\Rightarrow\; t_k = |\overline{x_k}| - \frac{\lambda}{2}$$

Pero como $t \geq 0$, entonces:

$$t_k(\lambda) = max\{|\overline{x_k}| - \frac{\lambda}{2}, 0\}$$

Y además como en el optimo se da que:

$$\sum_{i=1}^{d} t_i = 1 \;\Rightarrow\; \sum_{i=1}^{d} max\{|\overline{x_i}| - \frac{\lambda}{2}, 0\} = 1$$

Así para encontrar el optimo debemos encontrar $\lambda > 0$ tal que:

$$\sum_{i=1}^{d} max\{|\overline{x_i}| - \frac{\lambda}{2}, 0\} - 1 = 0$$

Ahora definiremos:

$$g(\lambda) = \sum_{i=1}^{d} max\{|\overline{x_i}| - \frac{\lambda}{2}, 0\} - 1$$

Y lo que uno esta buscando es que:

$$g(\lambda) = 0$$

Ahora notemos que $g(\lambda)$ es decreciente por ende, esto lo resolveremos por bisección.

Como sabemos que debe existir al menos un valor de t_i que sea positivo, entonces:

$$max_{i=1,\ldots,d} \left(|\overline{x_i}| - \frac{\lambda}{2}\right) > 0$$

$\Rightarrow$

$$max_{i=1,\ldots,d} \left(|\overline{x_i}|\right) > \frac{\lambda}{2}$$

$\Rightarrow$

$$2 \cdot max_{i=1,...,d}\left(|\overline{x_i}|\right) > \lambda$$

Así, nuestro método de bisección lo haremos en el intervalo:

$$[0, 2 \cdot max_{i=1,...,d}\left(|\overline{x_i}|\right)]$$

Así de esta forma encontramos la proyección en cada iteración y así se ejecuta finalmente el algoritmo. Pero como el algoritmo es de backtracking lo que haremos es corroborar que se cumpla que:

$$f(x^{t+1}) \leq f(x^t) + < \nabla f(x^t), x^{t+1} - x^t > + \frac{1}{2\eta_t}\|x^{t+1} - x^t\|_2^2$$

que es justamente el lema de descenso. Si no se cumple no se considera el x^{t+1} obtenido y se hace $\eta_t = \frac{\eta_t}{2}$.

Así finalmente el algoritmo nos queda como:

1. Sea $x^0 \in \mathcal{X} = \{x : \|x\|_1 \leq 1\}$ y $T \in \mathbb{N}$.
2. $\eta_t = 1$.
3. for $t = 0, ..., T-1$:

   a. Se realiza el calculo de $x^t - \eta_t \nabla f(x^t)$

   b. Se proyecta $x^t - \eta_t \nabla f(x^t)$ sobre la bola unitaria respecto a la norma 1, usando

   bisección.

   c. Se corrobora el lema de descenso.

   d. Si no se cumple, no se considera la iteración y se hace $\eta_t = \frac{\eta_t}{2}$

   e. Si se cumple, se considera la iteración y se sigue.
4. return x^T

---

b) Para esta sección notemos que el algoritmo es del tipo:

1. Sea $x^0 \in \mathcal{X}$ dado y T dado.

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

   $x^{t+1} = argmin_{x\in\mathcal{X}}\{f(x^t) + < \nabla f(x^t), x - x^t > + \frac{1}{2\eta_t}\|x - x^t\|_1^2\}$

3. return x^T

Así de esta manera decimos que:

$$x^{t+1} = \frac{x^t - \eta_t \nabla f(x^t)}{\|x^t - \eta_t \nabla f(x^t)\|_1}$$

De este modo, usando backtracking tenemos que:

1. Sea $x^0 \in \mathcal{X} = \{x : \|x\|_1 \leq 1\}$ y $T \in \mathbb{N}$.

2. $\eta_t = 1$.

3. for $t = 0, ..., T-1$:

   a. Se realiza el calculo de $x^{t+1} = \frac{x^t - \eta_t \nabla f(x^t)}{\|x^t - \eta_t \nabla f(x^t)\|_1}$

   b. Se corrobora el lema de descenso.

   c. Si no se cumple, no se considera la iteración y se hace $\eta_t = \frac{\eta_t}{2}$

   d. Si se cumple, se considera la iteración y se sigue.

4. return x^T

---

c) Notemos que el método de gradiente condicional es el siguiente:

1. Sea $x^0 \in \mathcal{X}$ dado y T dado.

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

$z^t = argmin_{z \in \mathcal{X}} < \nabla f(x^t), z >$

$x^{t+1} = x^t + \eta_t (z^t - x^t)$

con $\eta_t \in [0, 1]$

3. return x^T

Ahora lo que haremos es para este caso en especifico calcular el algoritmo (para luego implementarlo).

Notemos que:

$$< \nabla f(x^t), z > = \sum_{i=1}^{d} [\nabla f(x^t)]_i \cdot z_i$$

Y nosotros queremos minimizar esto sujeto a $\sum_{i=1}^{d} |z_i| \leq 1$

Lo que haremos será hacer un cambio de variable $z_i = t_i \cdot signo(z_i)$ con $t_i \geq 0$

De este modo, queda lo siguiente:

$$|z_i| = |signo(z_i)||t_i| = 1 \cdot t_i = t_i$$

De esta forma el problema anterior queda como sigue:

$$min \sum_{i=1}^{d} [\nabla f(x^t)]_i \cdot t_i \cdot signo(z_i)$$

$$s.a. \sum_{i=1}^{d} t_i \leq 1$$

$$t_1, t_2, ..., t_d \geq 0$$

Notemos que la solución de las variables $signo(z_i)$ es $signo(z_i^*) = -signo([\nabla f(x^t)]_i)$ $\forall i \in \{1, ..., d\}$ y lo demostraremos.

$$\sum_{i=1}^{d} [\nabla f(x^t)]_i \cdot t_i \cdot signo(z_i^*) = \sum_{i=1}^{d} [\nabla f(x^t)]_i \cdot t_i \cdot [-signo([\nabla f(x^t)]_i)] = -\sum_{i=1}^{d} signo([\nabla f(x^t)]_i) \cdot [\nabla f(x^t)]_i \cdot t_i$$

$$= \sum_{i=1}^{d} (-|[\nabla f(x^t)]_i|) \cdot t_i \leq \sum_{i=1}^{d} [\nabla f(x^t)]_i \cdot t_i \cdot signo(z_i)$$

$\Rightarrow$

$$\sum_{i=1}^{d} [\nabla f(x^t)]_i \cdot t_i \cdot signo(z_i^*) \leq \sum_{i=1}^{d} [\nabla f(x^t)]_i \cdot t_i \cdot signo(z_i)$$

$\forall signo(z_i) \in \{-1, 1\}$ y $\forall i \in \{1, ..., d\}$.

De esta forma el problema queda como sigue:

$$min - \sum_{i=1}^{d} |[\nabla f(x^t)]_i| \cdot t_i$$

$$s.a. \sum_{i=1}^{d} t_i \leq 1$$

$$t_1, t_2, ..., t_d \geq 0$$

Notemos que este es un problema lineal y sabemos que la solución se encuentra en los vértices. De este modo, mis posibles candidatos de solución son $t = e_i\ \forall i \in \{1, ..., d\}$ (e_i son los vectores canonicos en $\mathbb{R}^d$).

Ahora si evaluamos en la función objetivo en la posible solución propuesta $t = e_k$ obtenemos que:

$$- \sum_{i=1}^{d} |[\nabla f(x^t)]_i| \cdot t_i = - \left(\sum_{i=1:i\neq 0}^{d} |[\nabla f(x^t)]_i| \cdot t_i \right) - |[\nabla f(x^t)]_k| \cdot t_k =$$

$$- \left(\sum_{i=1:i\neq 0}^{d} |[\nabla f(x^t)]_i| \cdot 0 \right) - |[\nabla f(x^t)]_k| \cdot 1 = -|[\nabla f(x^t)]_k|$$

Así tenemos que el mínimo en el J tal que $-|[\nabla f(x^t)]_J| = min_{i=1,\ldots,d}(-|[\nabla f(x^t)]_i|)$ lo cual es lo mismo a que $|[\nabla f(x^t)]_J| = max_{i=1,\ldots,d}(|[\nabla f(x^t)]_i|)$.

De este modo, la solución esta dada por:

$$t = e_J \text{ con } J : |[\nabla f(x^t)]_J| = max_{i=1,\ldots,d}(|[\nabla f(x^t)]_i|)$$

Lo cual es equivalente a que:

$$t_k = 0\ \ \forall\ k\{1, ..., d\} : k \neq J \quad t_J = 1 \text{ con } J : |[\nabla f(x^t)]_J| = max_{i=1,\ldots,d}(|[\nabla f(x^t)]_i|)$$

Pero nuestra variable realmente es z. Ahora recordando que:

$$z_i = t_i \cdot signo(z_i)$$

Entonces tenemos que:

$$z_k = 0 \ \ \forall\ k \in \{1, ..., d\} : k \neq J, \quad z_J = signo(z_J) \text{ con } J : |[\nabla f(x^t)]_J| = max_{i=1,...,d}(|[\nabla f(x^t)]_i|)$$

Pero habíamos dicho que el óptimo de $signo(z_i)$ es tal que $signo(z_i) = -signo([\nabla f(x^t)]_i)$ $\forall i \in \{1, ..., d\}$.

De este modo:

$$z_k = 0 \ \ \forall\ k\{1, ..., d\} : k \neq J \quad z_J = -signo([\nabla f(x^t)]_J) \text{ con } J : |[\nabla f(x^t)]_J| = max_{i=1,...,d}(|[\nabla f(x^t)]_i|)$$

Así logramos encontrar el z^t y como tenemos que usar el paso propuesto en cátedra $\eta_t = \dfrac{2}{t+2}$, entonces el algoritmo nos queda como sigue:

1. Sea $x^0 \in \mathcal{X}$ dado y T dado.
2. for $t = 0, ..., T-1$:

   Buscamos J tal que: $|[\nabla f(x^t)]_J| = max_{i=1,...,d}(|[\nabla f(x^t)]_i|)$

   $z^t = -signo([\nabla f(x^t)]_J) \cdot e_J$

   $\eta_t = \dfrac{2}{t+2}$

   $x^{t+1} = x^t + \eta_t(z^t - x^t)$
3. return x^T

Análisis de resultados

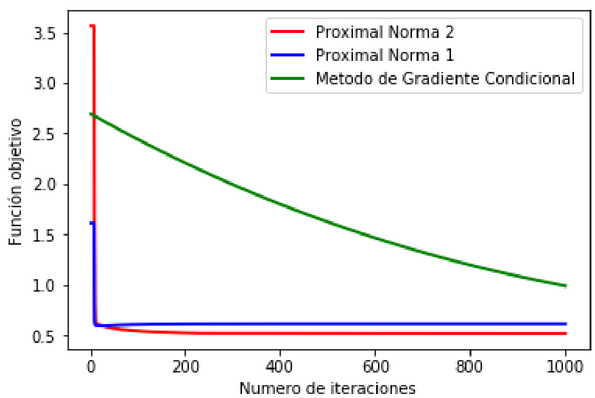
En esta sección lo que haremos será presentar los gráficos de los distintos algoritmos aplicados en los distintos casos (acompañado de sus respectivos análisis), además de un análisis de los costos computacionales de cada uno de los algoritmos en función del número de iteraciones.

Nota: Se uso la función $time.perfcounter()$ que Jupyter sugiría utilizar en vez de $time.clock()$.

Comenzaremos con los casos:

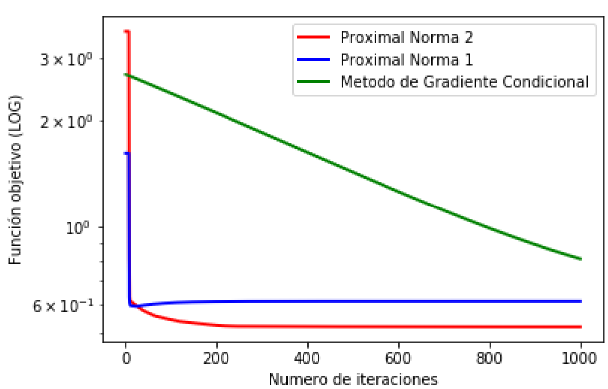
i) A10, b10:

a. Gráfico de la función objetivo $\frac{1}{2}\|Ax-b\|_2^2$.



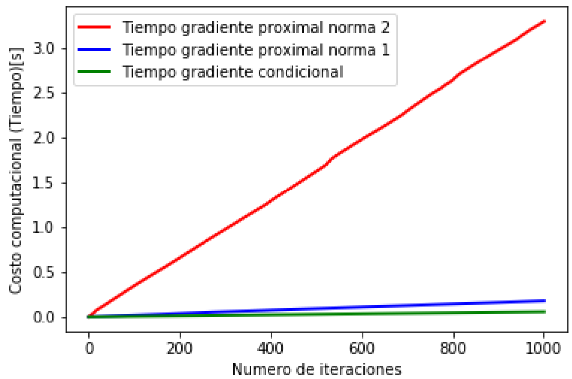
Comentarios: Como se observa el método del gradiente proximal con norma 2 es el que converge más rápido, mientras que el método del gradiente condicional converge más lentamente.

b. Gráfico de la función objetivo $\frac{1}{2}\|Ax-b\|_2^2$ en escala logarítmica.



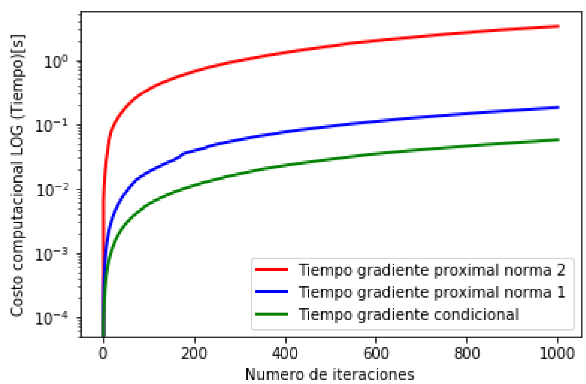
Comentarios: Al igual que en el gráfico anterior se observa el método del gradiente proximal con norma 2 es el que converge más rápido, mientras que el método del gradiente condicional converge más lentamente. El valor al que convergen las funciones es cercano a $6 \cdot 10^{-1}$.

c. Gráfico del costo computacional de los distintos algoritmos.



Comentarios: Como se observa el método del gradiente proximal con norma 1 tiene un costos computacionales que crecen linealmente, mientras que los otros dos métodos tienen costos computacionales que se mantienen relativamente constantes.

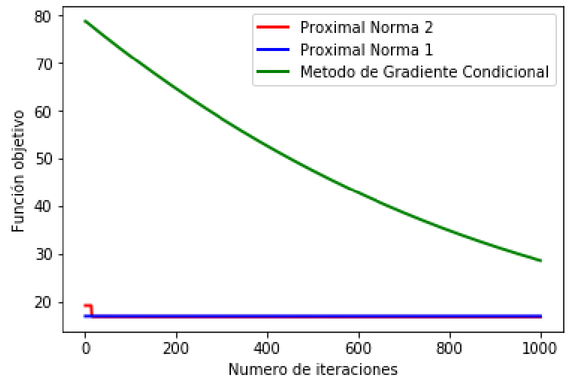
d. Gráfico del costo computacional de los distintos algoritmos en escala logaritmica.



Comentarios: Observando el grafico se puede corroborar lo que se dijo en el punto anterior, de que los costos del gradiente proximal crecen de manera lineal, porque en el gráfico en escala logaritmica se ve como una función logaritmica. Sin embargo, también acá se puede dilucidar que los otros dos métodos también crecen de manera lineal, pues en este grafico en escala logaritmica se ve que tienen forma de logaritmo, sin embargo, sus tasas de crecimiento son mucho menores que el método del gradiente proximal con norma 2. Es por esta razón que en el gráfico anterior los metodos del gradiente proximal con norma 1 y del gradiente restringido pareciesen que fueran constantes, pero la verdad es que crecen de manera lineal pero con una tasa muy pequeña.

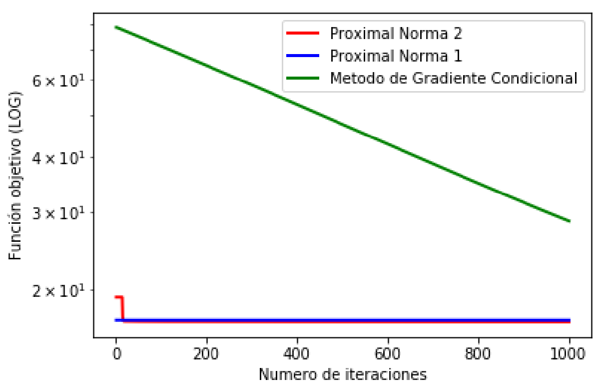
ii) A100, b100:

a. Gráfico de la función objetivo $\frac{1}{2}\|Ax - b\|_2^2$.



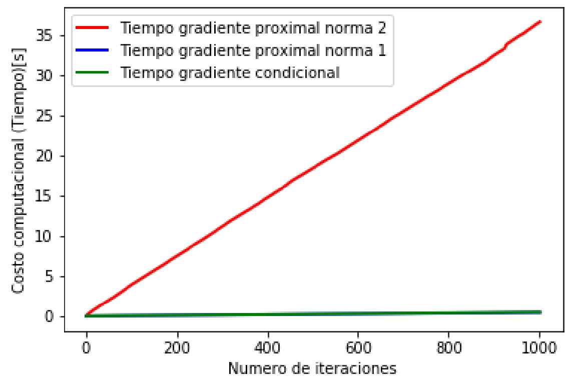
Comentarios: Como se observa el método del gradiente proximal con norma 1 y 2 convergen con una velocidad similar y el método del gradiente condicional converge de manera mas lenta.

b. Gráfico de la función objetivo $\frac{1}{2}\|Ax - b\|_2^2$ en escala logarítmica.



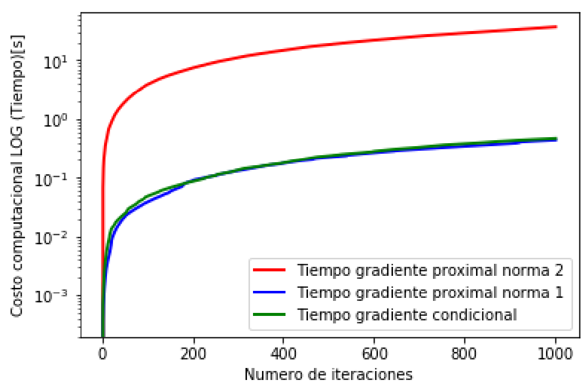
Comentarios: Al igual que en el gráfico anterior se observa que los métodos del gradiente proximal con norma 1 y 2 convergen más rapido que el método del gradiente condicional, el cual converge de manera exponencial, pues en escala logaritmica se ve que converge de manera lineal.

c. Gráfico del costo computacional de los distintos algoritmos.



Comentarios: Aquí se observa al igual que en el caso anterior que los costos computacionales del método del gradiente proximal con norma 2 crecen de manera lineal. Y en este caso se ve que los costos computaciones de los otros dos métodos son similares y parecen ser constantes.

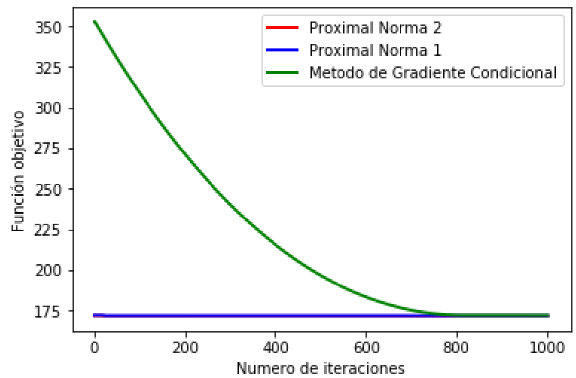
d. Gráfico del costo computacional de los distintos algoritmos en escala logaritmica.



Comentarios: Al igual que en el caso anterior, los 3 métodos crecen de manera lineal, porque en escala logaritmica todos parecen comportarse como logaritmo. La diferencia es que la tasa de crecimiento del método del gradiente proximal con norma 2 es mucho mayor que la de los otros dos métodos. Es por esta razón que en el gráfico anterior los metodos del gradiente proximal con norma 1 y del gradiente restringido pareciesen que fueran constantes, pero la verdad es que crecen de manera lineal pero con una tasa muy pequeña.

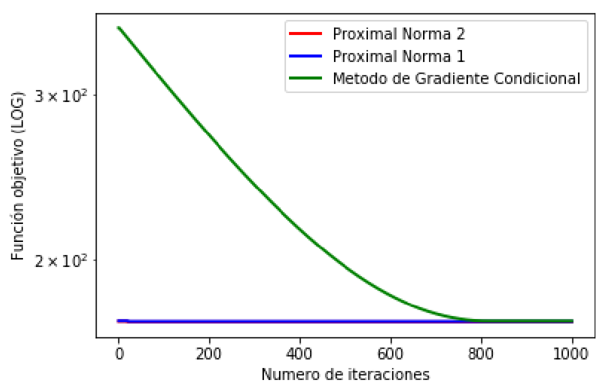
iii) A1000, b1000:

a. Gráfico de la función objetivo $\frac{1}{2}\|Ax - b\|_2^2$.



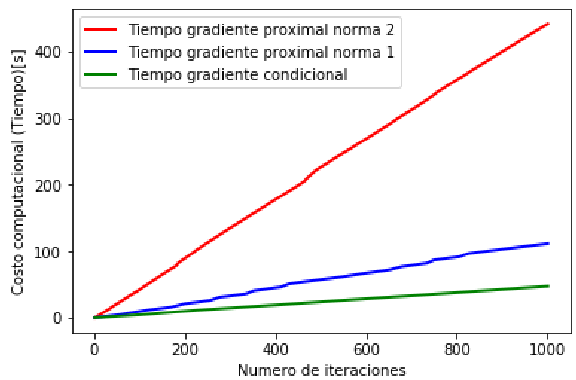
Comentarios: Como se observa el método del gradiente proximal con norma 1 y 2 convergen con una velocidad similar y el método del gradiente condicional converge de manera mas lenta.

b. Gráfico de la función objetivo $\frac{1}{2}\|Ax - b\|_2^2$ en escala logarítmica.



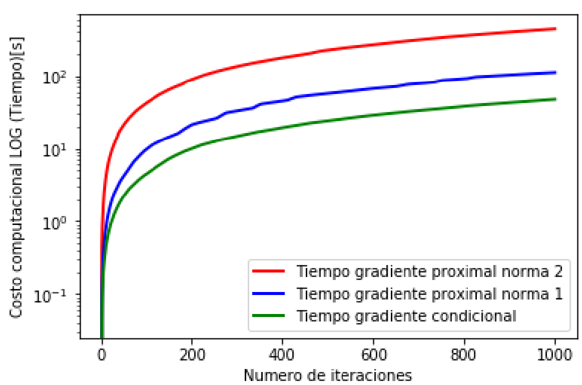
Comentarios: Se observa que los métodos del gradiente proximal con norma 1 y 2 convergen más rapido que el método del gradiente condicional, el cual converge de manera exponencial, pues en escala logaritmica se ve que converge de manera lineal.

c. Gráfico del costo computacional de los distintos algoritmos.



Comentarios: Notemoss que se observa que los costos de todos los algoritmos crecen de manera lineal. Pero la tasa de crecimiento del gradiente proximal con norma 2 es mayor que la tasa de crecimiento del gradiente proximal con norma 1 y esta tasa de crecimiento es mayor que la del método del gradiente condicional.

d. Gráfico del costo computacional de los distintos algoritmos en escala logaritmica.



Comentarios: Al igual que en el gráfico anterior se observa que los costos de todos los métodos crecen de manera lineal y la relación entre las tasas es la misma que la del gráfico anterior.

Observaciones generales:

Se observa que el método del gradiente proximal con norma 2 converge de una manera muy rápida sin embargo su costo computacional es bastante grande en comparación del método del gradiente condicional, el cual posee un menor costo computacional pero tiene una convergencia más lenta. Es así como a la hora de elegir que método utilizar es importante ver el trade off entre estos 2 aspectos además de ver que es lo que estamos priviligeando, ya sea un bajo costo computacional o una buena convergencia.