IMT3150: Optimización Avanzada 15 de abril de 2025

Tarea 1

Alumno: Vicente Opazo *Profesor: Cristóbal Guzmán*

Pregunta 1. *(a) Lo haremos a través de dos implicancias*

$\Rightarrow$: Consideremos un $(0, L_1)$ oráculo de f dado por $(\tilde{f}(x), \tilde{g}(x))$, veamos que entonces f es Gateaux derivable. Sea $x \in E$ arbitrario, veamos que f es Gateaux derivable para dicho punto. Primero notemos que de la definición de $(0, L_1)$ oráculo y considerando $y = x$ tenemos que

$$0 \leq f(x) - \left[\tilde{f}(x) + \langle \tilde{g}(x), x - x\rangle\right] \leq \frac{L_1}{2}\|x - x\|^2 \to 0 \leq f(x) - \tilde{f}(x) \leq 0 \to f(x) = \tilde{f}(x)$$

Nuevamente, de la definición de $(0, L_1)$ oráculo y considerando $y = x + \epsilon d$ se puede concluir que

$$\tilde{f}(x) + \langle \tilde{g}(x), (x + \epsilon d) - x\rangle \leq f(x + \epsilon d) \leq \tilde{f}(x) + \langle \tilde{g}(x), (x + \epsilon d) - x\rangle + \frac{L_1}{2}\|(x + \epsilon d) - x\|^2$$

Esto es

$$\tilde{f}(x) + \langle \tilde{g}(x), \epsilon d\rangle \leq f(x + \epsilon d) \leq \tilde{f}(x) + \langle \tilde{g}(x), \epsilon d\rangle + \frac{L_1}{2}\|\epsilon d\|^2$$

Restando por $\tilde{f}(x) = f(x)$ resulta en

$$\langle \tilde{g}(x), \epsilon d\rangle \leq f(x + \epsilon d) - f(x) \leq \langle \tilde{g}(x), \epsilon d\rangle + \frac{L_1}{2}\|\epsilon d\|^2$$

Dividiendo por ϵ

$$\langle \tilde{g}(x), d\rangle \leq \frac{f(x + \epsilon d) - f(x)}{\epsilon} \leq \langle \tilde{g}(x), d\rangle + \frac{\epsilon L_1}{2}\|d\|^2$$

Tomando el límite cuando ϵ tiende a 0

$$\langle \tilde{g}(x), d\rangle \leq \lim_{\epsilon \to 0} \frac{f(x + \epsilon d) - f(x)}{\epsilon} \leq \langle \tilde{g}(x), d\rangle$$

Y entonces

$$\lim_{\epsilon \to 0} \frac{f(x + \epsilon d) - f(x)}{\epsilon} = \langle \tilde{g}(x), d\rangle$$

Y por tanto la derivada en x es lineal, y el gradiente de x viene dado por $\nabla f(x) = \tilde{g}(x)$, demostrando así que f es Gateaux derivable, y que se cumple que $(\tilde{f}(x), \tilde{g}(x)) = (f(x), \nabla f(x))$. Así, con dicha equivalencia de la definición de $(0, L_1)$ oráculo tenemos que

$$f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L_1}{2}\|y - x\|^2 \qquad \forall x, y \in E$$

Lo cual corresponde a una caracterización de L_1-suave cuando f es convexa, demostrando así que es L_1-suave

$\Leftarrow$: Como f es L_1-suave, por una de las caracterizaciones vistas en clases tenemos que se cumple que

$$f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L_1}{2}\|y - x\|^2 \qquad \forall x, y \in E$$

Reordenando
$$f(y) - [f(x) + \langle \nabla f(x), y - x \rangle] \leq \frac{L_1}{2} \|y - x\|^2 \qquad \forall x, y \in E$$
De otra caracterización de L_1-suave tenemos que
$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L_1} \|\nabla f(x) - \nabla f(y)\|_*^2 \qquad \forall x, y \in E$$
Reordenando y notando que $\frac{1}{2L_1} \|\nabla f(x) - \nabla f(y)\|_*^2 \geq 0$ por ser una norma se obtiene que
$$f(y) - [f(x) + \langle \nabla f(x), y - x \rangle] \geq 0 \qquad \forall x, y \in E$$
Y entonces
$$0 \leq f(y) - [f(x) + \langle \nabla f(x), y - x \rangle] \leq \frac{L_1}{2} \|y - x\|^2 \qquad \forall x, y \in E$$
Y entonces por definición f posee un $(0, L_1)$ oráculo, ya que podemos considerar como dicho oráculo a $(\tilde{f}(x), \tilde{g}(x)) = (f(x), \nabla f(x))$, tal como queríamos demostrar.

(b) Veamos un ejemplo clásico, consideremos la función $f : \mathbb{R} \to \mathbb{R}$ dada por
$$f(x) = |x|$$
Esta función es convexa, puesto que
$$f(\lambda x + (1 - \lambda) y) = |\lambda x + (1 - \lambda) y| \leq \lambda |x| + (1 - \lambda)|y| = \lambda f(x) + (1 - \lambda) f(y) \qquad \forall x, y \in \mathbb{R}$$
Pero no es diferenciable en $x = 0$, puesto que
$$-1 = \frac{-\epsilon}{\epsilon} = \lim_{\epsilon \to 0^-} \frac{f(\epsilon) - f(0)}{\epsilon} \neq \lim_{\epsilon \to 0^+} \frac{f(\epsilon) - f(0)}{\epsilon} = \frac{\epsilon}{\epsilon} = 1$$
Y por tanto la función de derivada no es lineal, ya que no es continua.

Consideremos un oráculo de primer orden $(\tilde{f}(x), \tilde{g}(x))$ para f de la siguiente forma:
$$\tilde{f}(x) = f(x), \quad \tilde{g}(x) = \begin{cases} 1 & \textit{si } x > 0 \\ 0 & \textit{si } x = 0 \\ -1 & \textit{si } x < 0 \end{cases}$$

Para que esto defina un (δ, L)-oráculo de primer orden para f, queremos que se cumpla la siguiente desigualdad:
$$0 \leq f(y) - \left[\tilde{f}(x) + \langle \tilde{g}(x), y - x \rangle\right] \leq \frac{L}{2} \|y - x\|^2 + \delta, \quad \forall x, y \in \mathbb{R}$$

Para $x = 0$, se tiene $\tilde{f}(0) = 0$ y $\tilde{g}(0) = 0$, por lo que la desigualdad queda:
$$|y| \leq \frac{L}{2} y^2 + \delta$$

Esta desigualdad se cumple, por ejemplo, si tomamos $L = 1$ y $\delta = 2$, ya que claramente
$$|y| \leq \frac{1}{2} y^2 + 2, \quad \forall y \in \mathbb{R}$$

Para $x > 0$ la desigualdad queda

$$0 \leq f(y) - [f(x) + \langle 1, y - x\rangle] \leq \frac{1}{2}\|y - x\|^2 + 2, \quad \forall y \in \mathbb{R}$$

Esto es

$$0 \leq |y| - |x| + x - y \leq \frac{1}{2}\|y - x\|^2 + 2, \quad \forall y \in \mathbb{R}$$

Como $x > 0$

$$0 \leq |y| - y \leq \frac{1}{2}\|y - x\|^2 + 2, \quad \forall y \in \mathbb{R}$$

Si $y \geq 0$ entonces resulta $0 \leq 0 \leq \frac{1}{2}\|y - x\|^2 + 2$, lo cual es trivialmente cierto, y si $y < 0$ entonces resulta que $0 \leq 2y \leq \frac{1}{2}y^2 + 2 \leq \frac{1}{2}\|y - x\|^2 + 2$

Para $x < 0$ la desigualdad queda

$$0 \leq f(y) - [f(x) + \langle -1, y - x\rangle] \leq \frac{1}{2}\|y - x\|^2 + 2, \quad \forall y \in \mathbb{R}$$

Esto es

$$0 \leq |y| - |x| + y - x \leq \frac{1}{2}\|y - x\|^2 + 2, \quad \forall y \in \mathbb{R}$$

Como $x < 0$

$$0 \leq |y| - y \leq \frac{1}{2}\|y - x\|^2 + 2, \quad \forall y \in \mathbb{R}$$

Y procedemos como antes.

Por lo tanto, $(\tilde{f}, \tilde{g})$ define un $(1, 2)$-oráculo de primer orden para $f(x) = |x|$.

(c) Sea $(\tilde{f}(x), \tilde{g}(x))$ un (δ, L)-oráculo para f. De la primera desigualdad de (δ, L)-oráculo reemplazando $y = x$ se concluye que $\tilde{f}(x) \leq f(x)$

Por convexidad de f y derivabilidad Gâteaux

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle \geq \tilde{f}(x) + \langle \nabla f(x), y - x\rangle$$

Restando la desigualdad anterior a la segunda desigualdad correspondiente a (δ, L)-oráculo tenemos que

$$\langle \tilde{g}(x) - \nabla f(x), y - x\rangle \leq \frac{L}{2}\|y - x\|^2 + \delta$$

Sea $d \in E$ el vector que entrega su norma dual a $\tilde{g}(x) - \nabla f(x)$, esto es

$$\|\tilde{g}(x) - \nabla f(x)\|_* = \langle \tilde{g}(x) - \nabla f(x), d\rangle \quad \textit{con} \quad \|d\| = 1$$

Consideremos $y = x + t \cdot d$ para $t > 0$. Sustituyendo en la desigualdad anterior resulta que

$$\langle \tilde{g}(x) - \nabla f(x), x + t \cdot d - x\rangle \leq \frac{L}{2}\|x + t \cdot d - x\|^2 + \delta \Rightarrow \langle \tilde{g}(x) - \nabla f(x), t \cdot d\rangle \leq \frac{L}{2}\|t \cdot d\|^2 + \delta$$

Sacando las constantes, dado que d es un vector unitario y de la definición de norma dual se tiene que

$$t\|\tilde{g}(x) - \nabla f(x)\|_* \leq \frac{Lt^2}{2} + \delta \Rightarrow \|\tilde{g}(x) - \nabla f(x)\|_* \leq \frac{Lt}{2} + \frac{\delta}{t}$$

Para encontrar la mejor cota podemos optimizar el lado derecho para $t > 0$, derivando con respecto a t e igualando a 0 resulta que el óptimo se alcanza en $t = \sqrt{\frac{2\delta}{L}}$, y entonces

$$\|\tilde{g}(x) - \nabla f(x)\|_* \leq \frac{L}{2}\sqrt{\frac{2\delta}{L}} + \sqrt{\frac{L}{2\delta}}\delta = \sqrt{2\delta L}$$

Tal como queríamos demostrar.

(d) Sea $x \in E$. Notemos que de la primera desigualdad de la definición de oráculo se tiene que

$$\tilde{f}(x) - \langle \tilde{g}(x), x\rangle \leq f(y) - \langle \tilde{g}(x), y\rangle$$

Sea d el vector tal que $\|d\| = 1$ y $\|\tilde{g}(y) - \tilde{g}(x)\|_* = \langle \tilde{g}(y) - \tilde{g}(x), d\rangle$. Ahora consideremos la ecuación anterior pero reemplazando en y el valor $y - \frac{d\|\tilde{g}(y)-\tilde{g}(x)\|_*}{L}$, resulta que

$$\tilde{f}(x) - \langle \tilde{g}(x), x\rangle \leq f\left(y - \frac{d\|\tilde{g}(y) - \tilde{g}(x)\|_*}{L}\right) - \left\langle \tilde{g}(x), y - \frac{d\|\tilde{g}(y) - \tilde{g}(x)\|_*}{L}\right\rangle$$

Pero usando la segunda desigualdad de la definición de oráculo con $x := y$ e $y := y - \frac{d\|\tilde{g}(y)-\tilde{g}(x)\|_*}{L}$ se obtiene

$$\tilde{f}(x) - \langle \tilde{g}(x), x\rangle \leq \tilde{f}(y) + \left\langle \tilde{g}(y), -\frac{d\|\tilde{g}(y) - \tilde{g}(x)\|_*}{L}\right\rangle + \frac{L}{2}\left\|-\frac{d\|\tilde{g}(y) - \tilde{g}(x)\|_*}{L}\right\|^2 + \delta - \left\langle \tilde{g}(x), y - \frac{d\|\tilde{g}(y) - \tilde{g}(x)\|_*}{L}\right\rangle$$

Ordenando un poco

$$\tilde{f}(x) - \langle \tilde{g}(x), x\rangle \leq \tilde{f}(y) - \langle \tilde{g}(x), y\rangle + \left\langle \tilde{g}(y) - \tilde{g}(x), -\frac{d\|\tilde{g}(y) - \tilde{g}(x)\|_*}{L}\right\rangle + \frac{L}{2}\left\|-\frac{d\|\tilde{g}(y) - \tilde{g}(x)\|_*}{L}\right\|^2 + \delta$$

Usando la definición de d se tiene que

$$\tilde{f}(x) - \langle \tilde{g}(x), x\rangle \leq \tilde{f}(y) - \langle \tilde{g}(x), y\rangle - \frac{1}{L}\|\tilde{g}(y) - \tilde{g}(x)\|_*^2 + \frac{1}{2L}\|\tilde{g}(y) - \tilde{g}(x)\|_*^2 + \delta$$

Esto es

$$\frac{1}{2L}\|\tilde{g}(y) - \tilde{g}(x)\|_*^2 \leq \tilde{f}(y) - \tilde{f}(x) - \langle \tilde{g}(x), y - x\rangle + \delta$$

Pero de la primera desigualdad definición de oráculo usando $x = y$ se tiene $\tilde{f}(y) \leq f(y)$, y entonces

$$\frac{1}{2L}\|\tilde{g}(y) - \tilde{g}(x)\|_*^2 \leq f(y) - \tilde{f}(x) - \langle \tilde{g}(x), y - x\rangle + \delta$$

Tal como queríamos demostrar.

(e) De la pregunta (d) se tiene que

$$\|\tilde{g}(x) - \tilde{g}(y)\|_*^2 \leq 2L(f(y) - [\tilde{f}(x) + \langle \tilde{g}(x), y - x\rangle] + \delta)$$

Usando la definición de oráculo (la segunda desigualdad) tenemos que

$$\|\tilde{g}(x) - \tilde{g}(y)\|_*^2 \leq 2L \left(\frac{L}{2}\|y - x\|^2 + \delta + \delta \right)$$

Y desarrollando resulta que

$$\|\tilde{g}(x) - \tilde{g}(y)\|_* \leq \sqrt{L^2\|y - x\|^2 + 4\delta L}$$

Demostrando la primera estimación, para la segunda se tiene que por desigualdad triangular

$$\|\nabla f(x) - \nabla f(y)\|_* \leq \|\nabla f(x) - \tilde{g}(x)\|_* + \|\tilde{g}(x) - \tilde{g}(y)\|_* + \|\tilde{g}(y) - \nabla f(y)\|_*$$

Usando las cotas encontradas en (c) y (d) se tiene que

$$\|\nabla f(x) - \nabla f(y)\|_* \leq \sqrt{2\delta L} + \sqrt{L^2\|y - x\|^2 + 4\delta L} + \sqrt{2\delta L}$$

Y notando que $\sqrt{L^2\|y - x\|^2 + 4\delta L} \leq \sqrt{L^2\|y - x\|^2} + \sqrt{4\delta L}$ (puesto que al elevar al cuadrado el lado derecho resulta con un término adicional no negativo) se tiene que

$$\|\nabla f(x) - \nabla f(y)\|_* \leq \sqrt{2\delta L} + \sqrt{L^2\|y - x\|^2} + \sqrt{4\delta L} + \sqrt{2\delta L}$$

Y entonces, desarrollando resulta en

$$\|\nabla f(x) - \nabla f(y)\|_* \leq 2(\sqrt{2} + 1)\sqrt{\delta L} + L\|y - x\|$$

Tal como queríamos demostrar.

Pregunta 2. *(a) Veamos que f_λ es convexa. Sea $x_1, x_2 \in E$ y $\theta \in [0, 1]$. Sea $\varepsilon > 0$, y tomemos $u_1, u_2 \in E$ tales que*

$$f_\lambda(x_1) \geq f(u_1) + \frac{1}{2\lambda}\|x_1 - u_1\|^2 - \varepsilon, \quad f_\lambda(x_2) \geq f(u_2) + \frac{1}{2\lambda}\|x_2 - u_2\|^2 - \varepsilon$$

Llamemos $x = \theta x_1 + (1 - \theta)x_2$ $u = \theta u_1 + (1 - \theta)u_2$. Entonces tenemos por definición y luego por ser un punto sobre el que tomamos un ínfimo

$$f_\lambda(x) = \inf_{v \in E} \left\{ f(v) + \frac{1}{2\lambda}\|x - v\|^2 \right\} \leq f(u) + \frac{1}{2\lambda}\|x - u\|^2$$

Usando la convexidad de f y la desigualdad triangular (es decir, la convexidad de $\|\cdot\|^2$), tenemos que

$$f_\lambda(x) \leq \theta f(u_1) + (1 - \theta)f(u_2) + \frac{1}{2\lambda}\left(\theta\|x_1 - u_1\|^2 + (1 - \theta)\|x_2 - u_2\|^2\right)$$

Reorganizando

$$f_\lambda(x) \leq \theta \left[f(u_1) + \frac{1}{2\lambda}\|x_1 - u_1\|^2 \right] + (1 - \theta) \left[f(u_2) + \frac{1}{2\lambda}\|x_2 - u_2\|^2 \right]$$

Y por las desigualdades originales

$$f_\lambda(x) \leq \theta f(x_1) + (1 - \theta)f(x_2) + \epsilon$$

Como epsilon era arbitrario, haciéndolo tender a 0 podemos concluir que

$$f_\lambda(x) \leq \theta f(x_1) + (1 - \theta)f(x_2)$$

Demostrando así la convexidad de f_λ.

Para probar que $f_\lambda(x) \in \mathbb{R}$, primero notemos que $f_\lambda(x) \le f(x) + \frac{1}{2\lambda}\|x-x\|^2 = f(x)$, por ser un ínfimo, y por tanto no puede valer $+\infty$.

Para la otra cota notemos que

$$f(u) + \frac{1}{2\lambda}\|x-u\|^2 \ge \langle z, u\rangle + a + \frac{1}{2\lambda}\|x-u\|^2$$

Desarrollamos el término cuadrático

$$f(u) + \frac{1}{2\lambda}\|x-u\|^2 \ge \langle z, u\rangle + a + \frac{1}{2\lambda}\|u\|^2 - \frac{1}{\lambda}\langle x, u\rangle + \frac{1}{2\lambda}\|x\|^2$$

Agrupando términos

$$f(u) + \frac{1}{2\lambda}\|x-u\|^2 \ge \frac{1}{2\lambda}\|u\|^2 + \left\langle z - \frac{1}{\lambda}x, u\right\rangle + \left(a + \frac{1}{2\lambda}\|x\|^2\right)$$

Consideremos $c = z - \frac{1}{\lambda}x$. De lo que se tiene que completando el cuadrado

$$\frac{1}{2\lambda}\|u\|^2 + \langle c, u\rangle = \frac{1}{2\lambda}\left(\|u+\lambda c\|^2 - \lambda^2\|c\|^2\right)$$

Y entonces

$$f(u) + \frac{1}{2\lambda}\|x-u\|^2 \ge \frac{1}{2\lambda}\|u+\lambda c\|^2 + \left(a + \frac{1}{2\lambda}\|x\|^2 - \frac{\lambda}{2}\|c\|^2\right) \ge a + \frac{1}{2\lambda}\|x\|^2 - \frac{\lambda}{2}\left\|z - \frac{1}{\lambda}x\right\|^2$$

Así que la función está acotada inferiormente (ya que el lado derecho es constante) y entonces no se va a $-\infty$.

(b) La función $u \mapsto f(u) + \frac{1}{2\lambda}\|x-u\|^2$ es la suma de una función convexa con una función estrictamente convexa. Por tanto, es estrictamente convexa. Además, por lo visto en la parte (a) es coerciva, pues tiende a $+\infty$ cuando $\|u\| \to \infty$. Además, como el dominio tiene interior no vacío entonces la convexidad implica continuidad.

Así, por el teorema de Bolzano-Weiertrass extendido (creo que se llamaba así, me refiero al del curso de optimización inicial), el ínfimo es alcanzado en un único punto $u_x \in E$.

(c) Sea $x \in E$, tomando $u = x$ en la definición de f_λ, se tiene que

$$f_\lambda(x) \le f(x) + \frac{1}{2\lambda}\|x-x\|^2 = f(x)$$

Demostrando la primero cota. Ahora supongamos que f es L_0-Lipschitz. Sea u_x el punto donde se alcanza el mínimo en $f_\lambda(x)$, esto es

$$f_\lambda(x) = f(u_x) + \frac{1}{2\lambda}\|x-u_x\|^2$$

Usando la propiedad Lipschitz, y notando que $|f(x) - f(u_x)| = f(x) - f(u_x)$ ya que u_x es donde se alcanza el mínimo, resulta que

$$f(x) \le f(u_x) + L_0\|x-u_x\| = f_\lambda(x) - \frac{1}{2\lambda}\|x-u_x\|^2 + L_0\|x-u_x\|$$

Ahora estudiemos una cota superior para $-\frac{1}{2\lambda}\|x-u_x\|^2 + L_0\|x-u_x\|$*. Es una función cuadrática de* $\|x-u_x\|^2$ *y por tanto su máximo se alcanza cuando* $\|x-u_x\|^2 = \frac{-L_0}{2\cdot\frac{-1}{2\lambda}} = L_0\lambda$*, reemplazando resulta que*

$$-\frac{1}{2\lambda}\|x-u_x\|^2 + L_0\|x-u_x\|2 \leq -\frac{1}{2\lambda}(L_0\lambda)^2 + L_0\cdot(L_0\lambda) = \frac{-L_0^2\lambda}{2} + L_0^2\lambda = \frac{L_0^2\lambda}{2}$$

Y entonces

$$f(x) \leq f_\lambda(x) + \frac{L_0^2\lambda}{2} \leq f_\lambda(x) + 2\lambda L_0^2$$

Tal como queríamos demostrar

(d) Sea $\tilde{u}_\lambda(x)$ *el único punto donde se alcanza el mínimo, esto es*

$$f_\lambda(x) = f(\tilde{u}_\lambda(x)) + \frac{1}{2\lambda}\|x-\tilde{u}_\lambda(x)\|^2$$

Por la condición de Fermat para $\tilde{u}_\lambda(x)$ *en la función* $f(u) + \frac{1}{2\lambda}\|x-u\|^2$ *se tiene que*

$$\nabla f(\tilde{u}_\lambda(x)) + \frac{1}{\lambda}(\tilde{u}_\lambda(x) - x) = 0$$

Reordenando tenemos que

$$\nabla f(\tilde{u}_\lambda(x)) = \frac{1}{\lambda}(x - \tilde{u}_\lambda(x))$$

Obteniendo una de las igualdades que queríamos probar. Ahora notando que $f_\lambda(x) = \varphi(x, \tilde{u}_\lambda(x))$ *usamos la regla de la cadena para derivar con respecto a* x

$$\nabla f_\lambda(x) = \nabla_x\varphi(x, \tilde{u}_\lambda(x)) + \nabla_u\varphi(x, \tilde{u}_\lambda(x)) \cdot \nabla_x\tilde{u}_\lambda(x)$$

Pero recordando que por la condición de Fermat $\nabla_u\varphi(x, \tilde{u}_\lambda(x)) = 0$*, resulta que*

$$\nabla f_\lambda(x) = \nabla_x\varphi(x, \tilde{u}_\lambda(x)) = \frac{1}{\lambda}(x - \tilde{u}_\lambda(x))$$

Obteniendo la segunda igualdad que queríamos probar, esto es, podemos concluir que

$$\nabla f_\lambda(x) = \nabla f(\tilde{u}_\lambda(x)) = \frac{1}{\lambda}(x - \tilde{u}_\lambda(x))$$

Ahora veamos que f_λ *es* $\frac{1}{\lambda}$*-suave. Sean* $x, y \in E$*, entonces se tiene que por la convexidad de* f*, usando la monotonía del gradiente se tiene que*

$$\langle \nabla f(\tilde{u}_\lambda(x)) - \nabla f(\tilde{u}_\lambda(y)), \tilde{u}_\lambda(x) - \tilde{u}_\lambda(y)\rangle \geq 0$$

Reemplazando lo gradientes según lo calculado antes

$$\left\langle \frac{1}{\lambda}(x - \tilde{u}_\lambda(x)) - \frac{1}{\lambda}(y - \tilde{u}_\lambda(y)), \tilde{u}_\lambda(x) - \tilde{u}_\lambda(y) \right\rangle \geq 0$$

Reordenando

$$\left\langle \frac{1}{\lambda}(x - \tilde{u}_\lambda(x) - y + \tilde{u}_\lambda(y)), \tilde{u}_\lambda(x) - \tilde{u}_\lambda(y) \right\rangle \geq 0$$

Multiplicando por λ y reordenando resulta en

$$\langle x-y, \tilde{u}_\lambda(x)-\tilde{u}_\lambda(y)\rangle \geq \|\tilde{u}_\lambda(x)-\tilde{u}_\lambda(y)\|^2$$

Así, se tiene que

$$\|x-\tilde{u}_\lambda(x)-y+\tilde{u}_\lambda(y)\|^2 = \|x-y\|^2 - 2\langle x-y, \tilde{u}_\lambda(x)-\tilde{u}_\lambda(y)\rangle + \|\tilde{u}_\lambda(x)-\tilde{u}_\lambda(y)\|^2$$

Y por la desigualdad anterior

$$\|x-\tilde{u}_\lambda(x)-y+\tilde{u}_\lambda(y)\|^2 \leq \|x-y\|^2 \Rightarrow \|x-\tilde{u}_\lambda(x)-y+\tilde{u}_\lambda(y)\| \leq \|x-y\|$$

Ahora veamos por definición que

$$\|\nabla f_\lambda(x)-\nabla f_\lambda(y)\|_* = \left\|\frac{1}{\lambda}(x-\tilde{u}_\lambda(x)) - \frac{1}{\lambda}(y-\tilde{u}_\lambda(y))\right\|_* = \frac{1}{\lambda}\|x-\tilde{u}_\lambda(x)-y+\tilde{u}_\lambda(y)\|_*$$

Pero en espacios euclideanos la norma y la norma dual coinciden, por lo que

$$\|\nabla f_\lambda(x)-\nabla f_\lambda(y)\|_* = \frac{1}{\lambda}\|x-\tilde{u}_\lambda(x)-y+\tilde{u}_\lambda(y)\| \leq \frac{1}{\lambda}\|x-y\|$$

Tal como queríamos demostrar.

(e) Para que quede más clara la notación, a la única solución de (1) la voy a llamar $u^*(x)$. Por definición se tiene que

$$f_\lambda(y) = \varphi(y, u^*(y)) = f(u^*(y)) + \frac{1}{2\lambda}\|u^*(y)-y\|^2$$

Podemos tomar la diferencia de términos al cuadrado

$$\|u^*(y)-y\|^2 - \|u^*(y)-x\|^2 = \langle (u^*(y)-y)-(u^*(y)-x), (u^*(y)-y)+(u^*(y)-x)\rangle = \langle x-y, 2u^*(y)-x-y\rangle$$

Y entonces

$$f_\lambda(y) = f(u^*(y)) + \frac{1}{2\lambda}\left(\|u^*(y)-x\|^2 + \langle x-y, 2u^*(y)-x-y\rangle\right)$$

Esto es

$$f_\lambda(y) = \varphi(x, u^*(y)) + \frac{1}{2\lambda}\langle x-y, 2u^*(y)-x-y\rangle$$

De la definición de $\tilde{u}(x)$ se tiene que

$$\varphi(x,\tilde{u}(x)) - \varphi(x,u^*(y)) + \frac{1}{2\lambda}\|u^*(y)-\tilde{u}(x)\|^2 \leq \delta \Rightarrow \varphi(x,u^*(y)) \geq \varphi(x,\tilde{u}(x)) + \frac{1}{2\lambda}\|u^*(y)-\tilde{u}(x)\|^2 - \delta$$

Y entonces

$$f_\lambda(y) \geq \varphi(x,\tilde{u}(x)) + \frac{1}{2\lambda}\|u^*(y)-\tilde{u}(x)\|^2 - \delta + \frac{1}{2\lambda}\langle x-y, 2u^*(y)-x-y\rangle$$

Así

$$f_\lambda(y) \geq \varphi(x,\tilde{u}(x)) + \frac{1}{2\lambda}\|u^*(y)-\tilde{u}(x)\|^2 - \delta + \frac{1}{2\lambda}\langle x-y, (2\tilde{u}(x)-2x) + 2u^*(y) - 2\tilde{u}(x) - y + x\rangle$$

Usando la linealidad del producto interno

$$f_\lambda(y) \geq \varphi(x,\tilde{u}(x)) + \frac{1}{\lambda}\langle y-x, x-\tilde{u}(x)\rangle - \delta + \frac{1}{2\lambda}\|u^*(y)-\tilde{u}(x)\|^2 + \frac{1}{2\lambda}\langle x-y, 2u^*(y)-2\tilde{u}(x)-y+x\rangle$$

Pero

$$\|u^*(y)-\tilde{u}(x)\|^2+\langle x-y,x-y+2u^*(y)-2\tilde{u}(x)\rangle = \|u^*(y)-\tilde{u}(x)\|^2+2\langle x-y,u^*(y)-\tilde{u}(x)\rangle+\|x-y\|^2$$

Esto es

$$\|u^*(y)-\tilde{u}(x)\|^2+\langle x-y,x-y+2u^*(y)-2\tilde{u}(x)\rangle = \|(u^*(y)-\tilde{u}(x))+(x-y)\|^2 \geq 0$$

Y entonces

$$f_\lambda(y) \geq f(\tilde{u}(x)) + \frac{1}{2\lambda}\|x-\tilde{u}(x)\|^2 - \delta + \frac{1}{\lambda}\langle x-\tilde{u}(x), y-x\rangle$$

Demostrando la primera desigualdad de la definición de oráculo

Para la otra desigualdad notemos que por definición y por ser $u^*(y)$ mínimo se tiene que

$$f_\lambda(y) = f(u^*(y))+\frac{1}{2\lambda}\|y-u^*(y)\|^2 \leq f(\tilde{u}(x))+\frac{1}{2\lambda}\|y-\tilde{u}(x)\|^2 = f(\tilde{u}(x))+\frac{1}{2\lambda}\|(x-\tilde{u}(x))+(y-x)\|^2$$

Separando la norma convenientemente

$$f_\lambda(y) \leq f(\tilde{u}(x)) + \frac{1}{2\lambda}\|x-\tilde{u}(x)\|^2 + \frac{1}{2\lambda}\left(2\langle x-\tilde{u}(x), y-x\rangle + \langle y-x, y-x\rangle\right)$$

Esto es

$$f_\lambda(y) \leq \left(f(\tilde{u}(x)) + \frac{1}{2\lambda}\|x-\tilde{u}(x)\|^2 - \delta\right) + \frac{1}{\lambda}\langle x-\tilde{u}(x), y-x\rangle + \frac{1}{2\lambda}\|x-y\|^2 + \delta$$

Demostrando la segunda desigualdad de la definición de oráculo, y por tanto concluyendo que el par $(\tilde{f}(x), \tilde{g}(x))$ define un $\left(\delta, \frac{1}{\lambda}\right)$-oráculo de primer orden para f_λ, tal como queríamos demostrar.

Pregunta 3. *(a) Usando la actualización $x_{k+1} = x_k - \eta_k \tilde{g}_k(x_k)$ se tiene que*

$$\theta_{k+1} = \frac{1}{2}\|x_{k+1}-x^*\|^2 = \frac{1}{2}\|x_k - \eta_k\tilde{g}_k(x_k) - x^*\|^2 = \frac{1}{2}\langle x_k - \eta_k\tilde{g}_k(x_k) - x^*, x_k - \eta_k\tilde{g}_k(x_k) - x^*\rangle$$

Reordenando

$$\theta_{k+1} = \frac{1}{2}\|x_k - x^*\|^2 - \eta_k\langle\tilde{g}_k(x_k), x_k - x^*\rangle + \frac{\eta_k^2}{2}\|\tilde{g}_k(x_k)\|^2 = \theta_k - \eta_k\langle\tilde{g}_k(x_k), x_k - x^*\rangle + \frac{\eta_k^2}{2}\|\tilde{g}_k(x_k)\|^2$$

Usando la primera desigualdad de la definición del oráculo para $x := x_k$ e $y := x^*$ se tiene que

$$\langle\tilde{g}_k(x_k), x^* - x_k\rangle \leq f(x^*) - \tilde{f}_k(x_k) \Rightarrow -\langle\tilde{g}_k(x_k), x_k - x^*\rangle \leq f(x^*) - \tilde{f}_k(x_k) = v - \tilde{f}_k(x_k)$$

Reemplazando en la ecuación anterior

$$\theta_{k+1} \leq \theta_k + \eta_k(v - \tilde{f}_k(x_k)) + \frac{\eta_k^2}{2}\|\tilde{g}_k(x_k)\|^2$$

Ahora usando la segunda desigualdad de la definición de oráculo para $x := x_k$ e $y := x_{k+1}$ resulta que

$$f(x_{k+1}) - \tilde{f}(x_k) - \langle\tilde{g}_k(x_k), x_{k+1} - x_k\rangle \leq \frac{L_k}{2}\|x_{k+1} - x_k\|^2 + \delta_k$$

Usando que $x_{k+1} - x_k = -\eta_k \tilde{g}_k(x_k)$ se tiene que

$$f(x_{k+1}) - \tilde{f}(x_k) - \langle \tilde{g}_k(x_k), -\eta_k \tilde{g}_k(x_k) \rangle \leq \frac{L_k}{2} \| - \eta_k \tilde{g}_k(x_k)\|^2 + \delta_k$$

Esto es

$$f(x_{k+1}) - \tilde{f}(x_k) + \eta_k \|\tilde{g}_k(x)\|^2 \leq \frac{L_k \eta_k^2}{2} \|\tilde{g}_k(x_k)\|^2 + \delta_k$$

Pero como $\eta_k \leq \frac{1}{L_k}$

$$f(x_{k+1}) - \tilde{f}(x_k) + \eta_k \|\tilde{g}_k(x)\|^2 \leq \frac{\eta_k}{2} \|\tilde{g}_k(x_k)\|^2 + \delta_k$$

De lo que resulta

$$\frac{\eta_k}{2} \|\tilde{g}_k(x_k)\|^2 \leq -f(x_{k+1}) + \tilde{f}(x_k) + \delta_k$$

Esto es

$$\frac{\eta_k^2}{2} \|\tilde{g}_k(x_k)\|^2 \leq -\eta_k [f(x_{k+1}) - \tilde{f}(x_k) - \delta_k]$$

Y mezclandolo con la desigualdad de antes se obtiene

$$\theta_{k+1} \leq \theta_k + \eta_k (v - \tilde{f}_k(x_k)) - \eta_k [f(x_{k+1}) - \tilde{f}(x_k) - \delta_k]$$

Tal como queríamos demostrar.

(b) Si reordenamos la desigualdad de la parte anterior resulta en

$$\eta_k f(x_{k+1}) \leq \eta_k f(x^*) + \theta_k - \theta_{k+1} + \eta_k \delta_k$$

Sumando desde 0 hasta $K-1$ en ambos lados

$$\sum_{k=0}^{K-1} \eta_k f(x_{k+1}) \leq \sum_{k=0}^{K-1} (\eta_k f(x^*) + \theta_k - \theta_{k+1} + \eta_k \delta_k)$$

Cancelando la suma telescópica de los θ_k del lado derecho, resulta que

$$\sum_{k=0}^{K-1} \eta_k f(x_{k+1}) \leq \sum_{k=0}^{K-1} \eta_k f(x^*) + \sum_{k=0}^{K-1} \eta_k \delta_k + \theta_0 - \theta_K$$

Pero al ser $\theta_K \geq 0$ por ser una norma y sacando $f(x^*)$ de la sumatoria, se tiene que

$$\sum_{k=0}^{K-1} \eta_k f(x_{k+1}) \leq f(x^*) \sum_{k=0}^{K-1} \eta_k + \sum_{k=0}^{K-1} \eta_k \delta_k + \theta_0$$

Pero como f es convexa se tiene que

$$f(\bar{x}_K) \leq \frac{1}{\sum_{k=0}^{K-1} \eta_k} \sum_{k=0}^{K-1} \eta_k f(x_{k+1}) \Rightarrow f(\bar{x}_K) \sum_{k=0}^{K-1} \eta_k \leq \sum_{k=0}^{K-1} \eta_k f(x_{k+1})$$

Y entonces

$$f(\bar{x}_K) \sum_{k=0}^{K-1} \eta_k - f(x^*) \sum_{k=0}^{K-1} \eta_k \leq \sum_{k=0}^{K-1} \eta_k \delta_k + \theta_0$$

Y así, recordando que $\theta_0 = \frac{1}{2}\|x_0 - x^*\|^2$

$$f(\bar{x}_K) - f(x^*) \leq \frac{1}{\sum_{k=0}^{K-1} \eta_k} \left[\frac{1}{2}\|x_0 - x^*\|^2 + \sum_{k=0}^{K-1} \eta_k \delta_k \right]$$

Tal como queríamos demostrar. Ahora si tenemos que $\eta_k = \eta$, $\delta_k = \delta$ *y* $L_k = L$*, es decir, las constantes son fijas, el resultado anterior nos dice que*

$$f(\bar{x}_K) - f(x^*) \leq \frac{1}{\sum_{k=0}^{K-1} \eta} \left[\frac{1}{2}\|x_0 - x^*\|^2 + \sum_{k=0}^{K-1} \eta\delta \right] = \frac{1}{K\eta} \left[\frac{1}{2}\|x_0 - x^*\|^2 + K\eta\delta \right] = \frac{1}{2K\eta}\|x_0 - x^*\|^2 + \delta$$

Y recordando que $\eta \leq \frac{1}{L} \Rightarrow L \leq \frac{1}{\eta}$ *se concluye que*

$$f(\bar{x}_K) - f(x^*) \leq \frac{L\|x_0 - x^*\|^2}{2K} + \delta$$

Terminando la demostración.