Tarea 1

Pregunta 1. (30 %) Sea $(\mathbf{E}, \|\cdot\|)$ un e.v.n.f.d., y $f : \mathbf{E} \mapsto \mathbb{R}$ una función convexa. Dados $0 \leq \delta, L < \infty$, decimos que $\mathcal{O}_{\delta,L} : \mathbf{E} \mapsto \mathbb{R} \times \mathbf{E}$ es un (δ, L)-oráculo de primer orden para f si para todo $x \in \mathbf{E}$, si $\mathcal{O}(x) = (\tilde{f}(x), \tilde{g}(x))$ satisface

$$0 \leq f(y) - [\tilde{f}(x) + \langle \tilde{g}(x), y - x\rangle] \leq \frac{L}{2}\|y - x\|^2 + \delta.$$

(a) Pruebe que f posee un $(0, L_1)$-oráculo de primer orden ssi f es L_1-suave. Pruebe además que en tal caso, $(\tilde{f}(x), \tilde{g}(x)) = (f(x), \nabla f(x))$.

(b) De un ejemplo de una función convexa no-diferenciable junto a un (δ, L)-oráculo de primer orden para dicha función, con $\delta, L < +\infty$.

(c) Pruebe que si f es Gâteaux derivable y $(\tilde{f}(x), \tilde{g}(x))$ es un (δ, L)-oráculo de primer orden para f, entonces para todo $x \in \mathbf{E}$, $\|\tilde{g}(x) - \nabla f(x)\|_* \leq \sqrt{2\delta L}$.

(d) Pruebe que en el caso anterior, también se tiene que

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

(e) Concluya las siguientes estimaciones

$$\begin{aligned} \|\tilde{g}(x) - \tilde{g}(y)\|_* &\leq \sqrt{L^2\|x - y\|^2 + 4\delta L} \\ \|\nabla f(x) - \nabla f(y)\|_* &\leq 2(\sqrt{2} + 1)\sqrt{\delta L} + L\|x - y\|. \end{aligned}$$

Pregunta 2. (30 %) Sea $(\mathbf{E}, \|\cdot\|)$ un espacio Euclidiano, y sea $f : \mathbf{E} \mapsto \mathbb{R}$ una función convexa, no necesariamente Gâteaux derivable, pero con la siguiente propiedad: existen $z \in \mathbf{E}$ y $a \in \mathbb{R}$ tales que para todo $x \in \mathbf{E}$, $f(x) \geq \langle z, x\rangle + a$.* Dado $\lambda > 0$, consideramos la función

$$f_\lambda(x) = \inf_{u \in \mathbf{E}} \left\{ \varphi(x, u) := [f(u) + \frac{1}{2\lambda}\|x - u\|^2] \right\}. \tag{1}$$

(a) Pruebe que f_λ es convexa y que toma valores reales (es decir, el ínfimo no es $\pm\infty$),

(b) Pruebe que el ínfimo en (1) es alcanzado, y que la solución es única.

(c) Pruebe que $f_\lambda(x) \leq f(x)$, para todo $x \in \mathbf{E}$. Pruebe además que si f es L_0-Lipschitz, entonces $f(x) \leq f_\lambda(x) + 2\lambda L_0^2$, para todo $x \in \mathbf{E}$.

(d) Pruebe que si f es Gâteaux derivable** entonces f_λ es $1/\lambda$-suave. Pruebe además que $\nabla f_\lambda(x) = \nabla f(\tilde{u}_\lambda(x)) = \frac{1}{\lambda}[x - \tilde{u}_\lambda(x)]$, donde $\tilde{u}_\lambda(x)$ es la única solución de (1).

*Resulta ser que esta propiedad no es necesaria de agregar, pero ignoraremos este hecho.

**Al igual que antes, este supuesto es innecesario, pero por ahora usaremos este supuesto adicional.

(e) Suponga que, dado $x \in \mathbf{E}$, podemos calcular $\tilde{u}(x) \in \mathbf{E}$ tal que

$$\max_{u \in \mathbf{E}} \left\{ \varphi(x, \tilde{u}(x)) - \varphi(x, u) + \frac{1}{2\lambda}\|u - \tilde{u}(x)\|^2 \right\} \leq \delta. \tag{2}$$

Pruebe que el siguiente par define un $(\delta, 1/\lambda)$-oráculo de primer orden para f_λ,

$$\tilde{f}(x) = f(\tilde{u}(x)) + \frac{1}{2\lambda}\|x - \tilde{u}(x)\|^2 - \delta \tag{3}$$

$$\tilde{g}(x) = \frac{1}{\lambda}[x - \tilde{u}(x)] \tag{4}$$

Indicación. Use la fuerte-convexidad de φ con respecto a u.

Pregunta 3. (15 %) Sea $(\mathbf{E}, \|\cdot\|)$ un espacio Euclidiano. Consideramos ahora un problema de optimización convexa irrestricta

$$(P) \quad v = \min\{f(x) : x \in \mathbf{E}\},$$

con $\mathcal{S} := \arg\min\{f(x) : x \in \mathbf{E}\} \neq \emptyset$. Consideramos un análogo del método del gradiente con pasos $(\eta_k)_k$, donde sustuímos el oráculo de primer orden usual $(f(x), \nabla f(x))$ por un (δ_k, L_k)-oráculo $(\tilde{f}_k(\cdot), \tilde{g}_k(\cdot))$ (notar que permitimos que varíen los parámetros del oráculo en cada iteración); es decir

$$x_{k+1} = x_k - \eta_k \tilde{g}_k(x_k).$$

Supondremos además que $\eta_k \leq 1/L_k$, para todo k.

(a) Sea $x^* \in \mathcal{S}$. Pruebe que si $\theta_k := \frac{1}{2}\|x_k - x^*\|^2$, entonces

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

(b) Pruebe que $\bar{x}_K := \sum_{k=0}^{K-1} \frac{\eta_k}{\sum_{l=0}^{K-1}\eta_l} x_{k+1}$ satisface

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

En particular, concluya que para paso fijo $\eta_k \equiv \eta \leq 1/L$ (y donde $(\delta_k, L_k) \equiv (\delta, L)$)

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

Parte Computacional (25 %)

El objetivo de la parte computacional implementar un algoritmo con (δ, L)-oráculos de primer orden, en particular para modelos del tipo (1), donde la función a considerar corresponde a un problema de regresión lineal con penalización de exponente $1 < p < +\infty$

$$f(x) = \sum_{i=1}^{n} |\langle a_i, x\rangle - b_i|^p. \tag{5}$$

Note que esta función es convexa y Gâteaux derivable. La función objetivo será entregada en un formato matricial como sigue

$$A = \begin{bmatrix} b_1 & a_{11} & a_{12} & \dots & a_{1d} \\ b_2 & a_{21} & a_{22} & \dots & a_{2d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_m & a_{m1} & a_{m2} & \dots & a_{md} \end{bmatrix}.$$

(a) Implemente una subrutina `OraculoInexacto` (A, λ, x, δ), que dados datos en una matriz A, parámetros $\lambda, \delta > 0$ y $x \in \mathbb{R}^d$, implementa el oráculo inexacto $(\tilde{f}(x), \tilde{g}(x))$ dado en (3),(4) para f dada en (5). Para esto, implemente el método del gradiente con backtracking para diseñar una subrutina que garantice una solución para (2), dado un input (x, δ, λ) y un oráculo de primer orden para f.
Indicación. Note que por la Pregunta 2, el subproblema en (2) es fuertemente convexo. Si bien el objetivo (5) no es necesariamente suave (eso dependerá del exponente p), puede correr un método de gradiente con backtracking con criterio de parada $\|\nabla f(x_k)\|_2 \leq 10^{-4}$. Use resultados vistos en clase para poder comparar la garantía en norma de gradiente con la suboptimalidad en valor objetivo.

(b) Implemente el método de la Pregunta 3, que recibe como input $x_0 = 0 \in \mathbf{E}$ y una secuencia $(\delta_k, \lambda_k)_{k=0,\dots,K-1}$, y entrega $\bar{x}_K$.

(c) Ejecute su método en las instancias entregadas, y pruebe su método con las secuencias de parámetros (siempre use $\lambda = 10^{-3}$, $\eta_k = 1/L_k$ y $L_k = 1/\lambda$): $\delta_k \equiv 10^{-4}$ y $\delta_k = 1/\sqrt{k}$. Para las instancias unidimensionales, grafique tanto los datos como los modelos obtenidos. Para estos ejemplos, evalúe cualitativamente la calidad de los modelos obtenidos y compare el desempeño para los distintos valores de $p \in \{1{,}2,\ 2,\ 10\}$: para esto, use como criterios el error cuadrático medio sobre los datos provistos, así como el error cuadrático medio excluyendo a los outliers (para aquellos datos que contienen outliers: note que en los ejemplos provistos, el último 1 % de los datos observados corresponden a outliers).