PC4. ECOLE POLYTECHNIQUE. MAP 569. MACHINE LEARNING II.

EXERCISE 1 (PRELIMINARIES ON CONVEX ANALYSIS) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function and $\mathcal{C} \subset \mathbb{R}^d$ be a closed convex set. Condider the following optimisation problem

$$x^* \in \mathrm{argmin}_{x\in\mathcal{C}} f(x) \, .$$

Denote by $\pi_{\mathcal{C}} x = \mathrm{argmin}_{u\in\mathcal{C}} \|x-u\|^2$ the projection of x onto the convex set $\mathcal{C}$. The Projected Gradient Descent algorithm (with $\eta > 0$) is an algorithm to solve this problem:

$$\begin{aligned}
&\text{For } k = 1, \ldots, K-1 \, ,\\
&\qquad y_{k+1} = x_k - \eta \nabla f(x_k) \, ,\\
&\qquad x_{k+1} = \pi_{\mathcal{C}} y_{k+1} \, ,\\
&\text{Return} \quad f(x_K) \, .
\end{aligned}$$

1. Prove that for all $u \in \mathcal{C}$ and $0 < t < 1$, $\|z - (tu + (1-t)\pi_{\mathcal{C}} z)\|^2 \geqslant \|z - \pi_{\mathcal{C}} z\|^2$?
2. Deduce from the previous question that

$$\langle u - \pi_{\mathcal{C}} z, z - \pi_{\mathcal{C}} z\rangle \leqslant 0 \quad \text{and} \quad \|\pi_{\mathcal{C}} z - z\|^2 + \|u - \pi_{\mathcal{C}} z\|^2 \leqslant \|u - z\|^2.$$

3. Assume that f is differentiable and convex. For any $x, h \in \mathbb{R}^d$ and $t \in [0,1]$, define

$$F(t) = f(x + th) \, .$$

   Prove that $F(1) - F(0) \geqslant F'(0)$ and conclude that for all $x, y \in \mathbb{R}^d$,

$$f(y) - f(x) \geqslant \langle \nabla f(x), y - x\rangle.$$

EXERCISE 2 (CONVERGENCE RATES FOR LIPSCHITZ CONVEX FUNCTIONS) Assume that $\mathcal{C} \subset B(x_1, R)$ and let x^* be a minimizer of the optimization problem and define $\bar{x}_K = (x_1 + \ldots + x_K)/K$. In this section, we will prove that if $\|\nabla f(x)\| \leqslant L$ for all $x \in \mathcal{C}$, and $\eta = R/(L\sqrt{K})$, then

$$f(\bar{x}_K) - f(x^*) \leqslant \frac{LR}{\sqrt{K}} \, .$$

1. Using Exercice 1 Question 3, prove that

$$f(x_k) - f(x^*) \leqslant \frac{1}{\eta}\langle x_k - y_{k+1}, x_k - x^*\rangle = \frac{\eta}{2}\|\nabla f(x_k)\|^2 + \frac{1}{2\eta}\left(\|x_k - x^*\|^2 - \|y_{k+1} - x^*\|^2\right) .$$

2. Using Exercice 1 Question 2, prove that

$$\frac{1}{K}\sum_{k=1}^{K} f(x_k) - f(x^*) \leqslant \frac{\eta L^2}{2} + \frac{\|x_1 - x_*\|^2}{2\eta K}.$$

3. Conclude.

EXERCISE 3 (CONVERGENCE RATES FOR STRONGLY CONVEX FUNCTIONS) When the function f is strongly convex, then the PGD converges much faster. Assume that f is α-strongly convex:

$$f(x) - f(y) \leqslant \langle \nabla f(x), x - y\rangle - \frac{\alpha}{2}\|x - y\|^2 \tag{1}$$

and that ∇f is β-Lipschitz. The aim of this section is to prove that, for $\eta = 1/\beta$,

$$\|x_{K+1} - x^*\|^2 \leqslant \|x_1 - x^*\|^2 \, e^{-\rho K} \, ,$$

with $\rho = \alpha/\beta$. Define

$$g : x \mapsto \beta \left(x - \pi_{\mathcal{C}}(x - \frac{1}{\beta}\nabla f(x)) \right) .$$

The key of the proof is to obtain that, for all $(x, y) \in \mathcal{C}^2$,

$$f\left(x^+\right) - f(y) \leqslant \langle g(x), x - y\rangle - \frac{1}{2\beta}\|g(x)\|^2 - \frac{\alpha}{2}\|x - y\|^2 , \tag{2}$$

where $x^+ = \pi_{\mathcal{C}}(x - \beta^{-1}\nabla f(x))$.

1. Assume first that (2) holds. Prove the following (in)equalities:

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &= \|x_k - x^*\|^2 - \frac{2}{\beta}\langle g(x_k), x_k - x^*\rangle + \frac{1}{\beta^2}\|g(x_k)\|^2 , \\ &\leqslant (1-\rho)\|x_k - x^*\|^2 \leqslant e^{-\rho k}\|x_1 - x^*\|^2.\end{aligned}$$

2. It remains to prove (2). With the mean value theorem, prove that

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + t(y-x)), y - x\rangle \, dt \leqslant \langle \nabla f(x), y - x\rangle + \frac{\beta}{2}\|y - x\|^2 . \tag{3}$$

3. Remind that $x^+ = \pi_{\mathcal{C}}(x - \frac{1}{\beta}\nabla f(x))$. Using (1) and (3), check that

$$f(x^+) - f(y) \leqslant \langle \nabla f(x), x^+ - x\rangle + \frac{\beta}{2}\|x^+ - x\|^2 + \langle \nabla f(x), x - y\rangle - \frac{\alpha}{2}\|y - x\|^2 .$$

4. With Exercice 1 Question 2, prove that $\langle \nabla f(x), x^+ - y\rangle \leqslant \langle g(x), x^+ - y\rangle$ for all $y \in \mathcal{C}$.

5. Conclude that

$$\begin{aligned}f(x^+) - f(y) &\leqslant \langle g(x), x^+ - y\rangle + \frac{1}{2\beta}\|g(x)\|^2 - \frac{\alpha}{2}\|y - x\|^2 , \\ &= \langle g(x), x - y\rangle - \frac{1}{2\beta}\|g(x)\|^2 - \frac{\alpha}{2}\|y - x\|^2.\end{aligned}$$