Université Paris Saclay **MSc AI & MSc Data Science.**

TC2-Optimization Final Exam. December 19, 2024. Duration : 1h30. Faïcel Chamroukhi

Instructions : : Only pens are allowed (documents, calculators, phones, headphones, computers, tablets, are forbidden). It is forbidden to write with a pencil or a **red pen**. Your double exam sheet must include, in the designated area, your last name, first name, and signature. This designated area must be concealed by gluing. All your supplementary sheets must be numbered. The grading scale is provided for reference only.

* * *

Preliminaries : In the following let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable (and possibly twice-differentiable) function. If f is minimized iteratively, it is assumed that the initial point $x^{(0)}$ is close enough to $x^\star$ (i.e., $\|x^{(0)} - x^\star\|$ is small enough) where x^* is a critical point of f.

Theorem 1 : If f is μ-strongly convex and L-smooth then the sequence $(x_k)_{k\geq 0}$ generated by gradient descent iterates with a fixed step size $\alpha = \frac{1}{L}$ satisfies $f(x_t) - f(x^*) \leq \exp\left(-\frac{k\mu}{L}\right)(f(x_0) - f(x^*))$where κ is the condition number for $f(x)$.

Theorem 2 : If f is convex and L-smooth then the sequence $(x_k)_{k\geq 0}$ generated by gradient descent iterates with a fixed step size $0 \leq \alpha \leq \frac{1}{L}$ satisfies $f(x^{(k)}) - f(x^\star) \leq \frac{\|x^{(0)} - x^\star\|^2}{2\alpha k}$.

Exercice 1 *(2 ;2 ;2 ;2)* Consider $f(x) = \frac{1}{2}x^\top Ax$ where $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$.

1. Show whether $f(x)$ is μ-strongly convex (in that case determine the value of μ) or only convex.
2. Show that $f(x)$ is L-smooth and determine the value of L.
3. For this problem determine the convergence rate of gradient descent with step size $\alpha = \frac{1}{2}$ and comment on the result.
4. What happens if we have $A = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$?

Exercice 2 *(2 ;2 ;3 ;1)* Let $X \in \mathbb{R}^{n\times p}$ $(p < n)$, $y \in \mathbb{R}^n$ and $\lambda > 0$.

1. Prove that the following problem always admits a closed-form global solution :
$$\min_{\theta\in\mathbb{R}^p} \frac{1}{2}\|y - X\theta\|_2^2 + \frac{\lambda}{2}\|\theta\|_2^2.$$
2. Prove that for this problem the Newton algorithm converges globally and exactly in one iteration.
3. Now consider the following constrained minimization problem :
$$\min_{\theta\in\mathbb{R}^p} \frac{1}{2}\|y - X\theta\|_2^2 \quad \text{subject to} \quad \|\theta\|_2^2 \leq c$$
where c is a constant $c > 0$ representing a constraint on the norm of θ. Show that solving this constrained problem (2.) is equivalent to solving the unconstrained prblem (1.) with the following relation between the regularization parameter λ and the Lagrange multiplier μ :
$$\lambda = 2\mu.$$
4. Discuss the role of the multiplier μ (and consequently the role of the regularizer λ).

Exercice 3 *(2 ;2)* Consider minimizing the following problem iteratively :

$$\min_{\theta\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^\top\theta)^2,$$

where $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$ for $i = 1, \ldots, n$. Suppose n is very large, and the admissible computational cost per iteration is bounded by $O(m)$, where $m \ll n$. Propose a gradient algorithm to solve this problem efficiently. Justify your choice and discuss both its statistical and computational properties.