Université Paris Cité
UFR de Mathématiques et Informatique
45, rue des Saints-Pères, 75006, Paris.

Université Paris Cité

Master 1ère année, MMA, 2022-2023
Optimisation

Examen du 09/01/2023

Durée 2h. Aucun document n'est autorisé. Les exercices sont indépendants.

Exercice 1 (Cours)

1. Soit $U \subset \mathbb{R}^n$ un ensemble convexe non vide et $f : U \to \mathbb{R}$ une fonction convexe sur U. On suppose de plus que f admet en $x \in U$ un minimum local.
   a) Que peut-on alors dire de x pour f ?
   b) Prouver le résultat.
2. Soit $\Omega \subset \mathbb{R}^n$ un ouvert de $\mathbb{R}^n$ et $U \subset \Omega$ un sous-ensemble convexe non vide de Ω. Soit $f : \Omega \to \mathbb{R}$ une fonction différentiable sur Ω et on suppose f convexe sur U. Soit $x \in U$.
   a) Donner une condition nécessaire et suffisante pour que x soit un minimum global de f et faisant intervenir $\nabla f(x)$.
   b) Prouver cette CNS.
   c) Que devient la CNS si on a $U = \Omega$? *On n'attend pas de justification.*

Exercice 2

Soit $M, n \in \mathbb{N}\backslash\{0, 1\}$. Dans la suite, on considère $\mathbb{R}^M$ et $\mathbb{R}^n$ munis de leurs produits scalaires canoniques respectifs que l'on note (par abus de langage) $\langle \cdot, \cdot \rangle$ indépendamment de la dimension. Associés à ces produits scalaires, on utilise la notation unique (par abus de langage) $\|\cdot\|$ pour les normes euclidiennes.

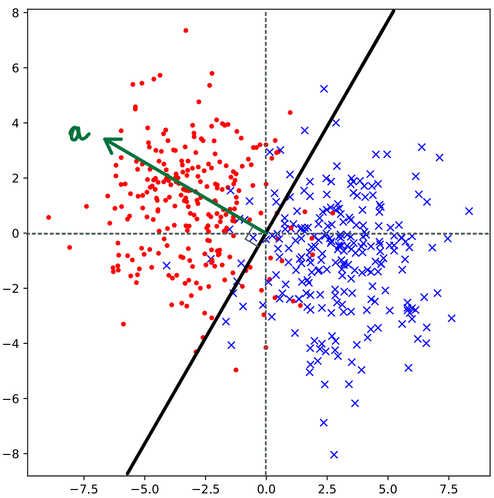

Figure 1. Exemple de points $x_i \in \mathbb{R}^2$. Ceux représentés par une croix, respectivement par un point, appartiennent à la classe -1, resp. 1. La droite représentée, notée $\mathcal{D}$, est $\{x \in \mathbb{R}^2 : \langle x, a \rangle = 0\}$. Vu le sens de a, tous les $x \in \mathbb{R}^2$ à gauche de $\mathcal{D}$ satisfont $\langle x, a \rangle > 0$, et à droite de $\mathcal{D}$, $\langle x, a \rangle < 0$. On associe à $f_a : x \in \mathbb{R}^n \mapsto \langle x, a \rangle$ (définissant en particulier $\mathcal{D}$), entièrement caractérisée par $a \in \mathbb{R}^2$, un coût $F(a)$. L'idée est de trouver un a de plus faible coût $F(a)$, en espérant que le F définie par le modèle permettent d'aboutir à un f_a effectuant une "bonne" classification.

On suppose donnée une séquence de M points $x_1, x_2, \ldots, x_M$ de $\mathbb{R}^n$, et pour chacun d'entre eux est associé une classe $y_i \in \{-1, 1\}$ pour $i \in \{1, \ldots, M\}$. Voir la figure 1 pour une illustration. On dit que la famille des couples, éléments de $\mathbb{R}^n \times \{-1, 1\}$, $((x_i, y_i))_{1 \leqslant i \leqslant M}$ forment les *données d'entraînement*.

On souhaite trouver un hyperplan (ou une droite dans le cas $n = 2$ comme sur la figure 1) passant par 0 et séparant, du mieux possible, ces points x_i entre ceux qui sont de la classe représentée par -1 et ceux de la classe représentée par 1.

Pour cela on considère la fonction suivante, pour $\lambda > 0$

$$F : a \in \mathbb{R}^n \mapsto \frac{1}{2M} \sum_{i=1}^{M} (1 - y_i \langle x_i, a \rangle)^2 + \frac{\lambda}{2} \|a\|^2 .$$

Et on s'intéresse au problème d'optimisation suivant

$$\inf_{a \in \mathbb{R}^n} F(a).$$

Partie I

1. Soit $a \in \mathbb{R}^n$ tel que la quantité $\frac{1}{2M} \sum_{i=1}^{M} (1 - y_i \langle x_i, a \rangle)^2$ est "petite". Pourquoi peut-on alors avoir confiance que la fonction $f_a : x \in \mathbb{R}^n \mapsto \langle x, a \rangle$ permette de séparer "correctement" les points x_i en fonction de leurs classes y_i ? Et comment cette séparation est-elle effectuée ?
2. Justifier que F est $\mathcal{C}^\infty$ sur $\mathbb{R}^n$.
3. Montrer que F admet un minimum global sur $\mathbb{R}^n$, que l'on notera $a_\lambda^\star \in \mathbb{R}^n$.
4. a) Montrer que pour tout $a \in \mathbb{R}^n$,

$$\nabla F(a) = \frac{1}{M} \sum_{i=1}^{M} (y_i \langle x_i, a \rangle - 1) y_i x_i + \lambda a.$$

   b) Déterminer pour tout $a \in \mathbb{R}^n$, $\nabla^2 F(a)$.
5. a) Montrer que F est strictement convexe.
   b) En déduire l'ensemble $\arg\min_{a \in \mathbb{R}^n} F(a)$.

Partie II

On va maintenant revisiter les résultats précédents en adoptant un point de vue matriciel. On définit le vecteur $\mathbb{1} = (1, \ldots, 1) \in \mathbb{R}^M$ et $Z \in \mathcal{M}_{n,M}(\mathbb{R})$ la matrice dont la i-ème colonne est donnée par le vecteur $y_i x_i = (y_i x_{i,1}, \ldots, y_i x_{i,n}) \in \mathbb{R}^n$.

6. Rappeler la définition d'une fonctionnelle quadratique $G : \mathbb{R}^n \to \mathbb{R}$.
7. a) Montrer que pour tout $a \in \mathbb{R}^n$, $F(a) = \frac{1}{2M} \left\| \mathbb{1} - Z^T a \right\|^2 + \frac{\lambda}{2} \|a\|^2$.
   b) En déduire que F est une fonctionnelle quadratique.
8. Donner alors, en utilisant les notations introduites dans cette partie, une expression de $\nabla F(a)$ et $\nabla^2 F(a)$, pout tout $a \in \mathbb{R}^n$.
9. En déduire une expression de $a_\lambda^\star$, puis $\lim_{\lambda \to +\infty} a_\lambda^\star$. Pourquoi est-ce logique d'obtenir cette limite sachant la manière dont est construite la fonction objectif F ?

Partie III

10. Justifier pourquoi on peut appliquer l'algorithme de descente de gradient à pas optimal à F.
11. Montrer que les hypothèses du théorème de convergence de l'algorithme de descente de gradient à pas optimal appliqué à F sont bien satisfaites.
12. Écrire en Python une fonction codant cet algorithme, prenant en argument un point initial $a_0 \in \mathbb{R}^n$, une tolérance ε pour la condition d'arrêt et des arguments supplémentaires que vous jugerez nécessaires et que vous définirez avant la fonction. Enfin cette fonction renverra le dernier itéré correspondant à la variable a.
13. Dans cette question uniquement on suppose $n = 2$. La trajectoire visible sur la figure 2 peut-elle correspondre à celle produite par l'algorithme de descente de gradient à pas optimal appliquée à F ? En plus de la justification, on en donnera une rapide preuve.
14. La figure 3 fournit une illustration de la convergence de l'algorithme de descente de gradient à pas optimal appliquée F. D'après le graphe, quelle type de convergence a-t-on pour la suite $(F(a^{(k)}))_{k \in \mathbb{N}}$, où $(a^{(k)})_{n \in \mathbb{N}}$ est la suite des itérés produites par l'algorithme en démarrant $a^{(0)} \in \mathbb{R}^n$.

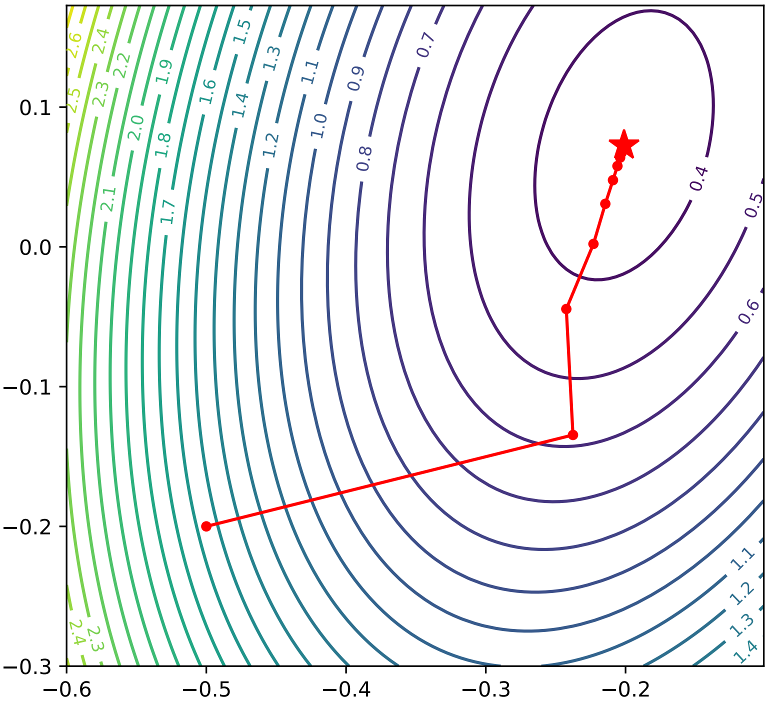

FIGURE 2. Les points correspondent aux éléments d'une suite $(a^{(k)})_{k\in\mathbb{N}}$, convergeant vers $a^\star_\lambda$ (représenté par l'étoile sur la figure), produite grâce à un algorithme itératif. Deux points reliés par un segment correspondent à deux itérés successifs de l'algorithme. Ainsi un segment donné a pour direction $a^{(k+1)} - a^{(k)}$, pour un certain $k \in \mathbb{N}$. En arrière plan sont représentées des lignes de niveau de F.

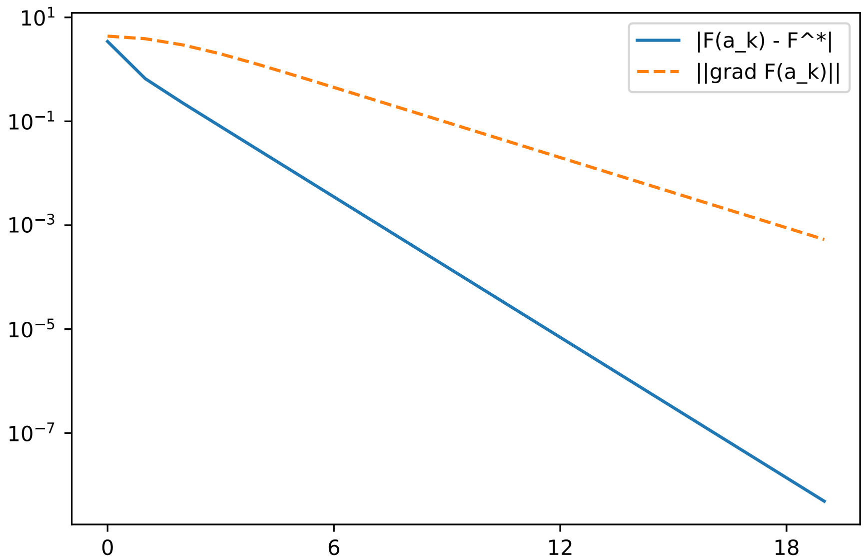

FIGURE 3. Illustration de la convergence de l'algorithme de descente de gradient à pas optimal appliquée à F.

Exercice 3

Soit $n \in \mathbb{N}^*$. Soit f une fonction $\mathcal{C}^2$ m-fortement convexe $(m > 0)$ sur $\mathbb{R}^n$ muni de sa norme euclidienne notée $\|\cdot\|$. De plus, on suppose que ∇f est L-Lipschitzienne ($L > 0$), i.e. pour tous $x, y \in \mathbb{R}^n$,

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L \|x - y\| .$$

Soit $x_0 \in \mathbb{R}^n$ et $\tau > 0$ fixé. On considère la suite $(x_k)_{k\in\mathbb{N}}$ définie par récurrence par

$$\forall k \in \mathbb{N}, \quad x_{k+1} = F(x_k),$$

où pour tout $x \in \mathbb{R}^n$, $F(x) = x - \tau\nabla f(x)$. On veut trouver une condition sur $\tau > 0$ de sorte que la suite $(x_k)_{k\in\mathbb{N}}$ converge.

1. Comment nommeriez-vous cette méthode produisant la suite $(x_k)_{k\in\mathbb{N}}$?
2. Étudier la convexité de f et montrer que f admet un unique minimum global, noté $x^\star$, sur $\mathbb{R}^n$.
3. Montrer que pour tous $x, y \in \mathbb{R}^n$,

$$\langle \nabla f(x) - \nabla f(y),\, x - y\rangle \geqslant m \|x - y\|^2 .$$

4. a) En développant pour tous $x, y \in \mathbb{R}^n$ la quantité $\|F(x) - F(y)\|^2$, en déduire que

$$\|F(x) - F(y)\|^2 \leqslant (1 + \tau^2L^2 - 2m\tau) \|x - y\|^2 .$$

b) En déduire que si $0 < \tau < \frac{2m}{L^2}$, alors il existe $\cancel{0 < \alpha_\tau < 1}$ $0 \leqslant \alpha_\tau < 1$ tel que pour tout $k \in \mathbb{N}^*$

$$\|x_{k+1} - x_k\| \leqslant \alpha_\tau \|x_k - x_{k-1}\| .$$

On suppose $0 < \tau < \frac{2m}{L^2}$ dans la suite.

c) En déduire que $(x_k)_{k\in\mathbb{N}}$ converge. On notera $\ell \in \mathbb{R}^n$ sa limite. *Indication : on cherchera à montrer que $(x_k)_{k\in\mathbb{N}}$ est une suite de Cauchy. On pourra commencer par majorer pour tout $k \in \mathbb{N}$, tout $p \in \mathbb{N}^*$,* $\|x_{k+p} - x_k\|$.

5. Montrer que la convergence de $(x_k)_{k\in\mathbb{N}}$ vers ℓ est linéaire.
6. Montrer que $\ell = x^\star$.
7. a) Soit $x, h \in \mathbb{R}^n$. En considérant $\psi : t \in \mathbb{R} \mapsto \langle \nabla f(x + th),\, h\rangle$ et en faisant deux développements de Taylor bien choisis, déduire que $\left\langle \nabla^2 f(x)h,\, h\right\rangle \leqslant L \|h\|^2$.

b) On pose $c = \frac{L}{m}$. Que représente cette quantité ?

c) Trouver la valeur optimale de τ, que l'on exprimera en fonction de c, garantissant la convergence la plus rapide de $(x_k)_{k\in\mathbb{N}}$ vers $x^\star$. Commenter la vitesse de convergence de $(x_k)_{k\in\mathbb{N}}$ lorsque c se trouve successivement dans ses deux régimes extrêmes.