Université de Paris - M1 MMA 2020-2021

Optimisation - Examen en présentiel du 04/01/2021

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

Exercice 1. Question de cours (2.5 points)

Enoncer et démontrer le résultat donnant une condition nécessaire de minimum local sur la matrice hessienne pour une fonction $f : \Omega \to \mathbb{R}$ deux fois différentiables sur un ouvert $\Omega \subset \mathbb{R}^n$.

Exercice 2. (5.5 points)

Soit f une fonction $\mathcal{C}^2$ m-fortement convexe ($m > 0$) sur $\mathbb{R}^n$, c'est-à-dire pour tous $x, h \in \mathbb{R}^n$,

$$m \|h\|_2^2 \leq \left\langle \nabla^2 f(x)h,\, h \right\rangle,$$

et telle que ∇f soit L-Lipschitzienne ($L > 0$), c'est-à-dire pour tous $x, y \in \mathbb{R}^n$,

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L \|x - y\|_2 .$$

Soient $x_0 \in \mathbb{R}^n$ et $\tau > 0$. 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 $x \in \mathbb{R}^n$, $F(x) = x - \tau \nabla f(x)$. C'est la méthode de descente de gradient à pas fixe. On veut trouver une condition sur $\tau > 0$ de sorte que la suite $(x_k)_{k\in\mathbb{N}}$ converge.

1. Ecrire une fonction Python *gradientPasFixe*, associée à cette méthode de descente, prenant en argument le gradient de f *gradf*, le pas de descente *tau*, un point de départ *x0*, une tolérance *tol*, un nombre maximal d'itérations *niter* et renvoyant le dernier itéré.

2. En utilisant un développement de Taylor approprié (on indiquera bien lequel), démontrer que pour tous $x, y \in \mathbb{R}^n$,
$$\langle x - y,\, \nabla f(x) - \nabla f(y) \rangle \geq m \|x - y\|_2^2 .$$

3. En développant pour tous $x, y \in \mathbb{R}^n$ $\|F(x) - F(y)\|_2^2$, en déduire que
$$\|F(x) - F(y)\|_2^2 \leq (1 + \tau^2 L^2 - 2m\tau) \|x - y\|_2^2 .$$

4. En déduire que si $0 < \tau < \frac{2m}{L^2}$, alors il existe $0 \leq c < 1$ tel que pour tout $k \geq 1$,
$$\|x_{k+1} - x_k\|_2 \leq c \|x_k - x_{k-1}\|_2 .$$

5. En déduire que $(x_k)_{k\in\mathbb{N}}$ converge vers un $x_* \in \mathbb{R}^n$. (*Indication : On cherchera à montrer que c'est une suite de Cauchy*).

6. Que représente ce x_* ? Justifier.

7. Quelle est la manière appropriée en Python pour illustrer la vitesse de convergence de $(x_k)_{k\in\mathbb{N}}$?

Exercice 3. (5.5 points)

Soient $n \in \mathbb{N}^*$. On définit pour tout $x \in \mathbb{R}^n$,

$$f(x) = \frac{1}{2} \|x - y\|_2^2 + \lambda \sqrt{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 + \varepsilon},$$

où $y \in \mathbb{R}^n$ est un vecteur fixé, $\lambda > 0$ et $\varepsilon > 0$.

1. Montrer que f peut s'écrire pour tout $x \in \mathbb{R}^n$,

   $$f(x) = \frac{1}{2} \|x - y\|_2^2 + \lambda \sqrt{\|Dx\|_2^2 + \varepsilon},$$

   avec $D \in \mathcal{M}_{n,n}(\mathbb{R})$ une matrice bien choisie.

2. Soient $\phi : \mathbb{R}_+ \to \mathbb{R}$ une fonction convexe et croissante, et $h : \mathbb{R}^n \to \mathbb{R}_+$ une fonction convexe. Montrer alors que $\phi \circ h$ est convexe.

3. Montrer que $g : x \in \mathbb{R}^n \mapsto \lambda \sqrt{\|Dx\|_2^2 + \varepsilon}$ est convexe. (*On pourra considérer la fonction* $t \in \mathbb{R}_+ \mapsto \lambda\sqrt{t^2 + \varepsilon}$).

4. Montrer que f est 1-fortement convexe.

5. Justifier que f admet un unique minimum global.

Exercice 4. (6.5 points)

On considère une fonction $f : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{C}^2$ et telle que il existe $0 < m \leq M$ satisfaisant

$$\forall x, h \in \mathbb{R}^n, \quad m \|h\|_2^2 \leq \left\langle \nabla^2 f(x) h,\, h \right\rangle \leq M \|h\|_2^2 .$$

Dans la suite on notera pour $v \in \mathbb{R}^n$, $\|v\|_1$ la norme ℓ_1 de v et $\|v\|_\infty$ la norme ℓ_∞ de v. La base canonique de $\mathbb{R}^n$ sera notée $(e_1, \ldots, e_n)$.

1. Soit $x \in \mathbb{R}^n$ fixé. On considère le problème de minimisation $\inf_{v \in \mathbb{R}^n, \|v\|_1 \leq 1} \langle \nabla f(x),\, v \rangle$.

   (a) Démontrer par un argument général que ce problème admet au moins une solution. (Remarque : une telle solution est une direction de plus forte pente en x pour la norme ℓ_1).

   (b) Démontrer que $v_* = -\operatorname{sign}(\partial_i f(x)) e_i$ où $i \in \{1, \ldots, n\}$ est défini par $|\partial_i f(x)| = \|\nabla f(x)\|_\infty$, est une solution du problème de minimisation. (*Indication : on commencera par déterminer une borne inférieure au problème de minimisation grâce à l'inégalité de Hölder suivante : pour tout* $v, v' \in \mathbb{R}^n$, $|\langle v,\, v' \rangle| \leq \|v\|_1 \|v'\|_\infty$.)

2. Soit $x^{(0)} \in \mathbb{R}^n$ un point quelconque. On définit par récurrence pour tout $k \in \mathbb{N}$, $x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)}$ où

   - $d^{(k)} = -\left\|\nabla f(x^{(k)})\right\|_\infty \operatorname{sign}(\partial_i f(x^{(k)})) e_i = -\partial_i f(x^{(k)}) e_i$ où i satisfait $|\partial_i f(x)| = \|\nabla f(x)\|_\infty$,
   - $t^{(k)} > 0$ satisfait $f(x^{(k)} + t^{(k)} d^{(k)}) = \min_{t>0} f(x^{(k)} + t d^{(k)})$.

   Soit $k \in \mathbb{N}$.

   (a) Justifier que $d^{(k)}$ est bien une direction de descente. ($t^{(k)}$ est donc supposé être le pas de descente optimal).

(b) Démontrer que pour tout $t > 0$

$$f(x^{(k)} + td^{(k)}) \leq f(x^{(k)}) - t \left\| \nabla f(x^{(k)}) \right\|_\infty^2 + \frac{t^2}{2} M \left\| \nabla f(x^{(k)}) \right\|_\infty^2 .$$

(*Indication : on pourra penser à utiliser un développement de Taylor-Maclaurin à l'ordre* 2).

(c) En déduire que $f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{1}{2M} \left\| \nabla f(x^{(k)}) \right\|_\infty^2$.

(d) En démontrant tout d'abord que $\|v\|_2^2 \leq n \|v\|_\infty^2$ pour tout $v \in \mathbb{R}^n$, puis en utilisant le fait que pour une fonction m-fortement convexe on a, $-\frac{1}{2m} \left\| \nabla f(x^{(k)}) \right\|_2^2 \leq f(x_*) - f(x^{(k)})$, démontrer que

$$f(x^{(k+1)}) - f(x_*) \leq (1 - \frac{m}{nM})(f(x^{(k)}) - f(x_*)),$$

où $x_* \in \mathbb{R}^n$ est l'unique minimum global de f.

(e) En déduire que $f(x_{k+1}) \underset{k \to +\infty}{\to} f(x_*)$. De quel type de convergence s'agit-il ?

(f) Comment évolue la convergence quand la dimension n augmente ? Est-ce logique ?