Concours d'Admission 2025 - Mathématiques (XUSR)

Écoles Normales Supérieures - École Polytechnique

Lundi 14 avril 2025

Durée : 4 heures
L'utilisation des calculatrices n'est pas autorisée pour cette épreuve.

Notations

On note $\mathcal{C}(\mathbb{R})$ l'ensemble des fonctions continues de $\mathbb{R}$ dans $\mathbb{R}$, et $\mathcal{C}^1(\mathbb{R})$ l'ensemble de celles qui sont dérivables sur $\mathbb{R}$ à dérivée continue. On dit qu'une fonction $g : \mathbb{R} \to \mathbb{R}$ est L-Lipschitzienne, où $L > 0$ est un nombre réel, si elle satisfait

$$\forall x, y \in \mathbb{R}, |g(x) - g(y)| \leq L|x - y|$$

On dit qu'une fonction $f : \mathbb{R} \to \mathbb{R}$ est convexe si elle satisfait

$$\forall x, y \in \mathbb{R}, \forall t \in [0, 1], f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

On pourra utiliser sans démonstration le fait qu'une fonction $f \in \mathcal{C}^1(\mathbb{R})$ est convexe si et seulement si sa dérivée f' est croissante.
Dans les parties I, II et III du sujet, la suite $(x_n)_{n\in\mathbb{N}}$ est toujours définie par la relation de récurrence

$$\forall n \in \mathbb{N}, x_{n+1} := x_n - \tau f'(x_n)$$

déterminée par un terme initial $x_0 \in \mathbb{R}$, un pas de temps $\tau > 0$, et une fonction $f \in \mathcal{C}^1(\mathbb{R})$.

Objectif de l'énoncé

L'objet de cette composition est d'établir certaines propriétés de l'algorithme de la descente de gradient et de ses variantes. Dans le cas général d'une fonction $f \in \mathcal{C}^1\left(\mathbb{R}^d\right)$ de plusieurs variables continûment différentiable, cette méthode s'écrit

$$\forall n \in \mathbb{N}, x_{n+1} = x_n - \tau \nabla f(x_n)$$

Le choix de la dimension $d = 1$ a été fait dans ce sujet (dernière partie exceptée) dans le but d'alléger les notations, cependant la plupart des résultats obtenus et des méthodes de preuve s'étendent naturellement en dimension arbitraire. Lorsque la fonction f possède de fortes propriétés (convexité, régularité, ...), la suite $(x_n)_{n\in\mathbb{N}}$ des itérées converge rapidement vers un minimiseur de f. Sous des hypothèses plus faibles, on peut parfois obtenir une convergence plus lente, ou bien avoir recours à une variante telle que la descente de gradient implicite. Comprendre le comportement fin de l'algorithme lorsque f a des propriétés plus faibles (par exemple f non-convexe) fait l'objet de recherches actuelles, et du champ d'investigation mathématique dit de l'optimisation numérique, dont les applications sont nombreuses (ingénierie, intelligence artificielle, etc).
À l'exception des préliminaires, toutes les parties sont indépendantes. Ne pas hésiter à admettre le résultat d'une question pour répondre aux suivantes.

Téléchargé gratuitement sur www.prepary.fr

Partie I : Préliminaires

Dans cette partie préliminaire, on établit d'abord l'existence d'un minimiseur, sous des hypothèses adéquates, puis une première propriété des itérées $(x_n)_{n\in\mathbb{N}}$ de la descente de gradient.

1. Soit $f \in \mathcal{C}(\mathbb{R})$ telle que
$$\lim_{x\to-\infty} f(x) = +\infty \quad \text{et} \quad \lim_{x\to+\infty} f(x) = +\infty$$
   (a) Montrer que l'ensemble $\{x \in \mathbb{R} \mid f(x) \leq f(0)\}$ est fermé et borné.
   (b) En déduire qu'il existe $x_* \in \mathbb{R}$ tel que $f(x_*) = \min\{f(x) \mid x \in \mathbb{R}\}$.
2. On suppose dans cette question que $f \in \mathcal{C}^1(\mathbb{R})$ est convexe, et que f' est L-Lipschitzienne, pour un certain $L > 0$.
   (a) Montrer que pour tous $x, y \in \mathbb{R}$
$$|f'(x) - f'(y)|^2 \leq L(x-y)\left(f'(x) - f'(y)\right)$$
   (b) Soient $x, y \in \mathbb{R}$, et soient $\hat{x} := x - \tau f'(x)$ et $\hat{y} := y - \tau f'(y)$. Montrer que
$$|\hat{x} - \hat{y}|^2 \leq |x-y|^2 - \tau(2-\tau L)(x-y)\left(f'(x) - f'(y)\right)$$
   (c) On suppose de plus que f admet un minimiseur x_*, et que $0 < \tau \leq 2/L$. Montrer que la suite $(|x_n - x_*|)_{n\in\mathbb{N}}$ est décroissante. (On rappelle que $(x_n)_{n\in\mathbb{N}}$ satisfait (2).)

Partie II : Convergence rapide, sous des hypothèses fortes

Dans cette partie, on montre que la suite $(x_n)_{n\in\mathbb{N}}$ définie par l'algorithme de descente de gradient converge rapidement vers un minimiseur de f, en parlant de convergence géométrique, en faisant des hypothèses fortes sur cette fonction. Commençons par l'étude d'un exemple.

1. Dans cette question seulement, on pose $f(x) := \frac{1}{2}Lx^2$ pour tout $x \in \mathbb{R}$, où $L > 0$ est fixé.
   (a) Montrer que $x_{n+1} = (1-\tau L)x_n$, puis exprimer directement x_n en fonction de x_0 et n.
   (b) On suppose $x_0 \neq 0$. Justifier que $x_n \to 0$ si et seulement si $0 < \tau < 2/L$.
2. Hypothèses : Dans la suite, on se donne $f \in \mathcal{C}^1(\mathbb{R})$ telle que f' est L-Lipschitzienne, avec $L > 0$, et on fixe τ tel que $0 < \tau \leq 2/L$. On suppose de plus que f est α-convexe, avec $\alpha > 0$, c'est-à-dire que
$$g(x) := f(x) - \frac{1}{2}\alpha x^2, \quad \text{est une fonction convexe sur } \mathbb{R}$$
   (a) Justifier que $f'(x) - \alpha x$ est une fonction croissante de $x \in \mathbb{R}$. En déduire que $\alpha \leq L$.
   (b) Montrer que $f(x) \geq f(0) + f'(0)x + \alpha x^2/2$ pour tout $x \in \mathbb{R}$. En déduire que f admet un minimiseur sur $\mathbb{R}$.
3. On note $x_* \in \mathbb{R}$ un minimiseur de f, dont l'existence vient d'être établie. Les hypothèses faites permettent d'établir que les itérées $(x_n)_{n\in\mathbb{N}}$ de la descente de gradient s'en rapprochent.
   (a) Montrer que $\alpha|x-y|^2 \leq \left(f'(x) - f'(y)\right)(x-y)$ pour tous $x, y \in \mathbb{R}$.
   (b) En déduire que pour tous $x, y \in \mathbb{R}$, en notant $\tilde{x} := x - \tau f'(x)$ et $\tilde{y} := y - \tau f'(y)$, on a
$$|\tilde{x} - \tilde{y}|^2 \leq |x-y|^2(1 - \alpha\tau(2 - L\tau))$$
   (c) On suppose $0 < \tau < 2/L$. Montrer que $|x_n - x_*| \leq \rho^n |x_0 - x_*|$, où ρ est une constante que l'on précisera, et telle que $0 \leq \rho < 1$.

Téléchargé gratuitement sur www.prepary.fr

Partie III : Convergence lente, sous des hypothèses faibles

Dans cette partie, on se passe de l'hypothèse très forte (4) utilisée précédemment. On montre que la suite $(x_n)_{n\in\mathbb{N}}$ des itérées de la descente de gradient converge, a priori assez lentement, vers un minimiseur de f, a priori non-unique. Commençons de nouveau par l'étude d'un exemple :

$$f(x) := \begin{cases} \frac{1}{3}x^3 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

1. Justifier que $f \in \mathcal{C}^1(\mathbb{R})$ et que f est convexe. Donner l'ensemble de ses minimiseurs.
2. On suppose dans cette question que $0 < x_0 < 1/\tau$.
   (a) Justifier que la suite $(x_n)_{n\in\mathbb{N}}$, définie par la relation de récurrence (2), est décroissante, à valeurs strictement positives, et satisfait $x_{n+1} = x_n(1 - \tau x_n)$ pour tout $n \in \mathbb{N}$. Justifier que $x_n \to 0$ lorsque $n \to \infty$.
   (b) Montrer que $1/x_{n+1} = 1/x_n + \tau/(1 - \tau x_n)$ pour $n \in \mathbb{N}$. En déduire que $x_n \leq x_0/(1 + n\tau x_0)$.
3. On suppose seulement $\tau > 0$. Montrer que pour tout $x_0 \in \mathbb{R}$, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers un minimiseur de f.

Hypothèses : On se donne dans la suite $f \in \mathcal{C}^1(\mathbb{R})$. On suppose que f est convexe, admet un minimiseur $x_* \in \mathbb{R}$, et que f' est L-Lipschitzienne. On suppose également que $0 < \tau < 2/L$.

1. Montrer que pour tous $x, y \in \mathbb{R}$
$$f(y) \geq f(x) + f'(x)(y - x)$$
   Indication : considérer un développement limité de (1) lorsque $t \to 0^+$.
2. Montrer que pour tous $x, y \in \mathbb{R}$
$$f(y) \leq f(x) + f'(x)(y - x) + \frac{L}{2}(y - x)^2$$
3. En déduire que pour tout $n \in \mathbb{N}$
$$f(x_{n+1}) \leq f(x_n) - \frac{\tau}{2}(2 - \tau L)\left|f'(x_n)\right|^2$$
4. En déduire que la suite $(f(x_n))_{n\in\mathbb{N}}$ est décroissante.

Dans les questions suivantes, on montre que l'algorithme du gradient converge en valeur, c'est-à-dire que la suite $(f(x_n))_{n\in\mathbb{N}}$ tend vers le minimum $f(x_*)$ de la fonction f.

1. Montrer que $0 \leq f(x) - f(x_*) \leq |x - x_*|\,|f'(x)|$ pour tout $x \in \mathbb{R}$.
2. Montrer que pour tout $n \in \mathbb{N}$, en supposant $x_0 \neq x_*$,
$$f(x_{n+1}) \leq f(x_n) - \frac{\tau}{2}(2 - \tau L)\frac{\left|f(x_n) - f(x_*)\right|^2}{\left|x_0 - x_*\right|^2}$$
   Indication : utiliser 2.c)
3. Soit $c > 0$, et soit $(a_n)_{n\in\mathbb{N}}$ une suite de nombres réels positifs telle que $a_{n+1} \leq a_n - c(a_n)^2$ pour tout $n \in \mathbb{N}$. Montrer $a_n \leq a_0/(1 + nca_0)$ pour tout $n \in \mathbb{N}$. Indication : adapter le raisonnement de la question 10.c)
4. En déduire une majoration de la suite de terme général $a_n := f(x_n) - f(x_*)$. Conclure que $f(x_n) \to f(x_*)$ lorsque $n \to \infty$.

Téléchargé gratuitement sur www.prepary.fr

On cherche maintenant à établir que la suite $(x_n)_{n\in\mathbb{N}}$ tend vers un minimiseur de f. On rappelle qu'on a noté $x_* \in \mathbb{R}$ un tel minimiseur, que l'on suppose exister, mais que celui-ci n'est pas forcément unique comme le montre l'exemple introductif.

1. Montrer que $\frac{\tau}{2}(2-\tau L)\sum_{0\leq i<n}|f'(x_i)|^2 \leq (f(x_0)-f(x_n))$ pour tout $n\in\mathbb{N}^*$. En déduire que $f'(x_n)\to 0$ lorsque $n\to\infty$.
2. Montrer que $(x_n)_{n\in\mathbb{N}}$ admet une sous-suite convergente. On note $\varphi : \mathbb{N}\to\mathbb{N}$ l'extractrice et x_{**} la limite correspondante, de sorte que $x_{\varphi(n)}\to x_{**}$ lorsque $n\to\infty$. Indication. On pourra utiliser sans démonstration le théorème de Bolzano-Weierstrass : de toute suite dans $\mathbb{R}$ bornée, on peut extraire une sous-suite convergente.
3. Montrer que $f'(x_{**})=0$.
4. En déduire que x_{**} est un minimiseur de f, puis que $|x_n-x_{**}|\to 0$ lorsque $n\to\infty$.

Partie IV : Descente de gradient proximale

Une limitation de l'algorithme de descente de gradient est que la fonction à minimiser doit être (continûment) dérivable. Dans cette partie on considère une généralisation de cet algorithme, qui se passe de cette hypothèse. Le lien avec la descente de gradient, dans sa variante implicite, est fait dans la question 21.

Hypothèses : On considère une fonction convexe $f\in\mathcal{C}(\mathbb{R})$, admettant un minimiseur $x_*\in\mathbb{R}$. Soit également $\tau>0$.

1. Montrer que la fonction $F_{x_0}(x) := \frac{1}{2}|x-x_0|^2+\tau f(x)$ admet un unique minimiseur sur $\mathbb{R}$, que l'on notera $p_f(x_0)$. Indication : On pourra considérer des minimiseurs x_1 et x_2 de F_{x_0}, et remarquer que
$$\left|\frac{1}{2}(x_1+x_2)-x_0\right|^2 < \frac{1}{2}|x_1-x_0|^2+\frac{1}{2}|x_2-x_0|^2 \quad \text{si } x_1\neq x_2$$
2. Montrer que $x_0\in\mathbb{R}$ est un minimiseur de f si et seulement si $p_f(x_0)=x_0$. Indication : considérer la quantité $F_{x_0}((1-t)x_0+tx_*)$ lorsque $t\to 0^+$.
3. Dans cette question seulement, on suppose que $f\in\mathcal{C}^1(\mathbb{R})$. Montrer que $x_1 := p_f(x_0)$ satisfait
$$x_1 = x_0-\tau f'(x_1)$$

Dans toute la suite, étant donné $x_0\in\mathbb{R}$, on définit la suite récurrente $(x_n)_{n\in\mathbb{N}}$ par
$$x_{n+1} := p_f(x_n)$$

Illustrons le comportement de cette suite sur un exemple.

1. Dans cette question seulement, on pose $f(x) := |x|$ pour tout $x\in\mathbb{R}$.
   (a) Montrer que
   $$p_f(x) = \begin{cases} x-\tau & \text{si } x\geq\tau \\ x+\tau & \text{si } x\leq -\tau \\ 0 & \text{sinon.}\end{cases}$$
   (b) En déduire que $x_n\to 0$ lorsque $n\to\infty$, quels que soient $x_0\in\mathbb{R}$ et $\tau>0$.

Les questions suivantes étudient les différences entre les termes successifs de la suite $(x_n)_{n\in\mathbb{N}}$, désormais définie par (5). (Attention : (2) n'a plus cours.)

Téléchargé gratuitement sur www.prepary.fr

1. Montrer que $\frac{1}{2}\left|x_1-x_0\right|^2+\tau f\left(x_1\right) \leq \tau f\left(x_0\right)$. En déduire que pour tous entiers $N>M \geq 0$

$$\frac{1}{2} \sum_{M<n \leq N}\left|x_n-x_{n-1}\right|^2 \leq \tau\left(f\left(x_M\right)-f\left(x_N\right)\right)$$

En déduire que $\left|x_{n+1}-x_n\right| \rightarrow 0$ lorsque $n \rightarrow \infty$.

2. Montrer que pour tous $M, N \in \mathbb{N}$

$$\left|x_N-x_M\right| \leq \sqrt{2 \tau|N-M|} \sqrt{\left|f\left(x_M\right)-f\left(x_N\right)\right|}$$

Montrons maintenant que la fonction p_f définissant notre suite récurrente est 1-Lipschitzienne.

1. Soient $x \in \mathbb{R}$, et $\tilde{x}:=p_f(x)$. Montrer que pour tous $v \in \mathbb{R}$ et $t \in \mathbb{R}$

$$\tau f(\tilde{x})+\frac{1}{2}|\tilde{x}-x|^2 \leq \tau f(\tilde{x}+t v)+\frac{1}{2}|\tilde{x}+t v-x|^2$$

Soient également $y \in \mathbb{R}$, et $\tilde{y}:=p_f(y)$. En déduire que

$$2 \tau(f(\tilde{x})+f(\tilde{y})-f(\tilde{x}+t v)-f(\tilde{y}-t v)) \leq|\tilde{x}+t v-x|^2+|\tilde{y}-t v-y|^2-|\tilde{x}-x|^2-|\tilde{y}-y|^2$$

2. Montrer que le membre de droite dans l'inégalité (6) admet le développement limité $2 t v(\tilde{x}-x+y-\tilde{y})$ lorsque $t \rightarrow 0$.
3. On choisit $v:=\tilde{y}-\tilde{x}$ dans l'inégalité (6). Montrer que le membre de gauche est positif pour tout $t \in[0,1]$. En déduire que

$$|\tilde{x}-\tilde{y}|^2 \leq(x-y)(\tilde{x}-\tilde{y})$$

4. Montrer que $\left|p_f(x)-p_f(y)\right| \leq|x-y|$ pour tous $x, y \in \mathbb{R}$. En déduire que la suite $\left(\left|x_n-x_*\right|\right)_{n \in \mathbb{N}}$ est décroissante.
5. On obtient comme dans la question 18.a) l'existence d'une sous-suite convergente $x_{\varphi(n)} \rightarrow x_{**}$ lorsque $n \rightarrow \infty$, où $\varphi: \mathbb{N} \rightarrow \mathbb{N}$ est strictement croissante et $x_* \in \mathbb{R}$.
6. Montrer que $x_{\varphi(n)+1} \rightarrow x_{**}$ lorsque $n \rightarrow \infty$, puis en déduire que $p_f\left(x_{**}\right)=x_{**}$.
7. Conclure que x_{**} est un minimiseur de f, et que $x_n \rightarrow x_{**}$ lorsque $n \rightarrow \infty$.

Partie V : Optimisation sur la boule unité

Dans cette partie, on étudie la variante dite projetée de l'algorithme descente de gradient, qui vise à minimiser une fonction de plusieurs variables, restreinte à la boule unité. On fixe un entier $d \in \mathbb{N}^*$, et on munit $\mathbb{R}^d$ du produit scalaire usuel noté $\langle\cdot, \cdot\rangle$ et de la norme euclidienne associée $\|\cdot\|$. On note $C:=\left\{x \in \mathbb{R}^d \mid\|x\| \leq 1\right\}$ la boule unité fermée de $\mathbb{R}^d$, et on se donne $f \in \mathcal{C}^1\left(\mathbb{R}^d\right)$.

1. Montrer que f admet un minimiseur sur C, que l'on note x_* dans les questions suivantes.
2. On suppose dans cette question que $\left\|x_*\right\|<1$. Montrer que $\nabla f\left(x_*\right)=0$.
3. On suppose dans cette question que $\left\|x_*\right\|=1$. L'objectif est de montrer que

$$\exists \lambda \geq 0, \nabla f\left(x_*\right)=-\lambda x_*$$

   (a) Soient $x, y \in \mathbb{R}^d$ tels que $x \neq y$ et $\|x\|=\|y\|=1$. Montrer que $\langle x, v\rangle>0$ et $\langle y, v\rangle<0$, où $v:=x-y$.

Téléchargé gratuitement sur www.prepary.fr

(b) On suppose par l'absurde que (7) n'est pas satisfaite. Montrer qu'il existe $v \in \mathbb{R}^d$ tel que $\langle v, \nabla f(x_*)\rangle > 0$ et $\langle v, x_*\rangle > 0$. En déduire une contradiction et conclure. Indication : considérer les quantités $f(x_* - tv)$ et $\|x_* - tv\|^2$, dans la limite $t \to 0^+$.

Dans la suite de cette section, M désigne une matrice réelle symétrique de taille $d \times d$ non nulle telle que

$$\forall x \in \mathbb{R}^d, \langle x, Mx\rangle \geq 0$$

Les vecteurs de $\mathbb{R}^d$ sont ici considérés comme des vecteurs colonnes. On définit la fonction f de $\mathbb{R}^d$ dans $\mathbb{R}$ par

$$f(x) := -\frac{1}{2}\langle x, Mx\rangle$$

1. Montrer que $\nabla f(x) = -Mx$, pour tout $x \in \mathbb{R}^d$.
2. Décrire l'ensemble des minimiseurs de f sur C.

Étant donné $x_0 \in \mathbb{R}^d$, on définit une suite récurrente par

$$x_{n+1} := P_C\left(x_n - \tau \nabla f\left(x_n\right)\right), \quad \text{avec} \quad P_C(x) := \begin{cases} x & \text{si } \|x\| \leq 1 \\ x/\|x\| & \text{sinon.} \end{cases}$$

Cette suite correspond à l'algorithme de la descente de gradient projetée.

1. On suppose dans cette question que $\|x_0\| \geq 1$.
   (a) Montrer que
   $$\forall n \in \mathbb{N}\backslash\{0\}, x_n = \frac{\left(\mathrm{I}_d + \tau M\right)^n x_0}{\left\|\left(\mathrm{I}_d + \tau M\right)^n x_0\right\|}$$
   (b) Calculer $\lim_{n\to\infty} x_n$. Indication. Décomposer $x_0 = \sum_{1\leq i\leq d} \alpha_i e_i$ dans une base orthonormée de vecteurs propres $(e_1, \cdots, e_d)$, associés aux valeurs propres $\lambda_1, \cdots, \lambda_d$ de M. Introduire l'ensemble d'indices $I := \{i \in [\![1, d]\!] \mid \alpha_i \neq 0\}$, la valeur propre $\lambda := \max_{i\in I} \lambda_i$, et le vecteur $x_0' := \sum_{i\in I'} \alpha_i e_i$ où $I' := \{i \in I \mid \lambda_i = \lambda\}$.
2. Comment la suite se comporte-t-elle lorsque $\|x_0\| < 1$?
3. Montrer qu'il existe un hyperplan $H \subset \mathbb{R}^d$ tel que, pour tout $x_0 \in \mathbb{R}^d\backslash H$, on a $\lim_{n\to\infty} f(x_n) = \min\{f(x) \mid x \in C\}$.

Téléchargé gratuitement sur www.prepary.fr