Agrégation Interne de Mathématiques
Thierry Champion
séances du 30 septembre et du 14 octobre 2016

Optimisation, méthode du gradient

On étudie le problème d'optimisation

$$\min\{f(x) : x \in \mathbb{R}\} \tag{P}$$

où f est une fonction de classe $\mathcal{C}^2$ sur $\mathbb{R}$ et atteignant son minimum sur $\mathbb{R}$. On notera $\hat{x}$ une solution de (P), une telle solution étant appelée **minimiseur** de f.

La méthode du gradient, sous sa forme "continue", est le système dynamique suivant :

$$\begin{cases} X(0) = x_0, \quad x_0 \text{ fixé} \\ \forall t \geq 0, \quad X'(t) = -f'(X(t)) \end{cases} \tag{M_C}$$

qu'on peut discrétiser sous forme explicite (M_E) ou implicite (M_I) de la manière suivante :

$$\begin{cases} X_0 = x_0,\ x_0 \text{ fixé} \\ \forall n \geq 0, \quad X_{n+1} = X_n - \delta\, f'(X_n) \end{cases} \tag{M_E}$$

ou

$$\begin{cases} X_0 = x_0,\ x_0 \text{ fixé} \\ \forall n \geq 0, \quad X_{n+1} = X_n - \delta\, f'(X_{n+1}) \end{cases} \tag{M_I}$$

où $\delta > 0$ est un paramètre fixé. On se propose d'étudier la convergence de ces méthodes vers un **point critique** de f, i.e. une solution de $f'(x) = 0$. On remarquera que $f'(\hat{x}) = 0$ pour tout minimiseur de f.

Le but de ce problème est de montrer, sous certaines conditions, que $X(t)$ (respectivement X_n) converge vers un minimiseur de f lorsque $t \to +\infty$ (resp. $n \to +\infty$).

1 Méthode du gradient, cas continu

On étudie la convergence de la méthode (M_C) : dans cette partie, la fonction $t \mapsto X(t)$ est une solution de (M_C). On fera de plus l'hypothèse que

$$\forall x, y \in \mathbb{R}, \qquad (f'(x) - f'(y))(x - y) \geq a_f\,(x - y)^2 \tag{1}$$

pour un réel $a_f \geq 0$ fixé.

1. Résoudre (M_C) dans le cas où $f(x) = (x-1)^2$ et $x_0 = 0$. Montrer que la limite de $X(t)$ quand $t \to +\infty$ est l'unique minimiseur de f.

2. Donner une condition nécessaire et suffisante sur f'' pour que (1) soit vérifiée.

3. Démontrer que tout point critique de f est un minimiseur de f.
   Dans le cas $a_f > 0$, démontrer qu'un tel minimiseur est unique.

4. On suppose dans cette question que $a_f > 0$. On définit la fonction $h(t) = (X(t) - \hat{x})^2$. En étudiant les variations de h, démontrer que $h(t) \to 0$ quand $t \to +\infty$. *On pourra admettre le fait suivant, valable pour tout réel α (lemme de Gronwall):*

$$\textit{si } h'(t) \leq \alpha h(t) \textit{ pour tout } t \in \mathbb{R}_+ \textit{ alors } h(t) \leq h(t_0) e^{\alpha(t-t_0)} \textit{ pour tout } t \geq t_0 \geq 0.$$

   Montrer que la limite de $X(t)$ quand $t \to +\infty$ est l'unique minimiseur de f.

5. On suppose désormais que $a_f = 0$.

   (a) En étudiant la fonction h, démontrer que la fonction $X(t)$ est bornée.

   (b) Etudier les variations de la fonction $\phi(t) = f(X(t))$, et en déduire que l'intégrale $\displaystyle\int_0^{+\infty} (f'(X(t)))^2 dt$ est convergente.

   (c) Démontrer que la fonction $(f'(X(t)))^2$ est décroissante et tend vers 0 quand $t \to +\infty$. En déduire que les points d'accumulation de $X(t)$ quand $t \to +\infty$ sont des minimiseurs de f.

   (d) En utilisant la fonction h pour un point d'accumulation $\hat{x}$ fixé de $X(t)$, en déduire que $X(t)$ converge vers ce minimiseur de f.

2 Méthode du gradient, cas discret

On fera également dans cette partie l'hypothèse (1) sur f.

1. Dans le cas où $f(x) = (x-1)^2$ et $x_0 = 0$, étudier la convergence de la suite $Y_n = X_n - 1$ en fonction de la valeur du paramètre $\delta > 0$ pour les méthodes (M_E) et (M_I).

2. On s'intéresse dans cette question à la méthode explicite (M_E) et on suppose que $a_f > 0$ et que f' est L-Lipschitzienne :

$$\forall x, y \in \mathbb{R}, \qquad |f'(x) - f'(y)| \leq L\,|x - y|. \tag{2}$$

   Démontrer que

$$\forall n, \qquad (X_{n+1} - \hat{x})^2 \;\leq\; (1 - 2a_f\delta + L^2\delta^2)(X_n - \hat{x})^2$$

   et en déduire les valeurs de δ pour lesquelles la convergence de la suite $(X_n)_n$ vers $\hat{x}$ est linéaire. Quel est alors le choix de δ assurant le meilleur taux de convergence ?
   Rappel : on parle de convergence linéaire de $(X_n)_n$ vers $\hat{x}$ si $\displaystyle\limsup_{n\to+\infty} \frac{|X_{n+1} - \hat{x}|}{|X_n - \hat{x}|} = c < 1$ où c est le taux de convergence; plus c est petit, plus la convergence est rapide.

3. On s'intéresse dans cette question à la méthode implicite (M_I) et on suppose $a_f \geq 0$.

   (a) Pour $n \geq 0$ fixé, démontrer que la fonction $g : x \mapsto f(x) + \frac{1}{2\delta}(x - X_n)^2$ vérifie la condition (1) pour une valeur $a_g > 0$. En déduire que X_{n+1} est l'unique minimiseur de g.

   (b) Déduire de la question précédente que la série $\sum (X_{n+1} - X_n)^2$ est convergente et que $f'(X_n) \to 0$ quand $n \to +\infty$. En déduire que les points d'accumulation de (X_n) sont des minimiseurs de f.

   (c) En étudiant la suite $(X_{n+1} - \hat{x})^2$, démontrer que la suite $|X_{n+1} - \hat{x}|$ est décroissante.

   (d) En choisissant pour $\hat{x}$ un minimiseur de f adéquat, démontrer que la suite $(X_n)_n$ converge vers un minimiseur de f.

3 Méthode du gradient dans $\mathbb{R}^d$, cas discret

Dans la suite f est une fonction de classe $\mathcal{C}^2$ sur $\mathbb{R}^d$ à valeurs dans $\mathbb{R}$ et atteignant son minimum sur $\mathbb{R}^d$ en au moins un point $\hat{x}$. On s'intéresse à la méthode du gradient explicite (M_{E,R^d}) qui s'écrit :

$$\begin{cases} X_0 = x_0,\ x_0 \text{ fixé} \\ \forall n \geq 0, \quad X_{n+1} = X_n - \delta\,\nabla f(X_n) \end{cases} \tag{M_{E,R^d}}$$

où $\delta > 0$ est un paramètre fixé et $\nabla f(x) = (\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_d}(x))$ est le gradient de f au point x.

On étudiera *exclusivement* le cas particulier

$$f(x) = \frac{1}{2}\|Ax - b\|^2$$

pour une matrice $A \in \mathbb{R}^{d\times d}$ et un vecteur $b \in \mathbb{R}^d$ fixés, où $\|.\|$ est la norme euclidienne. Ici, les vecteurs x et b sont représentés par des matrices colonnes (ce qui donne un sens au produit Ax). Pour tous $x, y \in \mathbb{R}^d$ on notera $x \cdot y$ le produit scalaire usuel de x et y, et on rappelle que $(Bx)\cdot y = x \cdot ({}^tBy)$ où ${}^tB$ est la transposée de la matrice $B \in \mathbb{R}^{d\times d}$.

1. Soit $x, u \in \mathbb{R}^d$ deux vecteurs fixés.

   (a) Calculer la dérivée en 0 de fonction numérique $t \mapsto f(x + tu)$. En déduire que $\nabla f(x) = {}^tA\,(Ax - b)$.

   (b) Expliciter la méthode (M_{E,R^d}) pour la fonction f lorsque $A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$ et $b = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$. Dans ce cas, vérifier que le minimum de f sur $\mathbb{R}^2$ est 0 et donner l'ensemble des minimiseurs de f.

2. Dans cette question on s'intéresse à la généralisation de l'hypothèse (1) à ce cadre :

$$\forall x, y \in \mathbb{R}^d, \qquad (\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq a_f\,\|x - y\|^2 \tag{3}$$

   (a) Soit $B \in \mathbb{R}^{d\times d}$ une matrice symétrique, démontrer que

$$\forall x \in \mathbb{R}^d, \qquad \lambda_1(B)\,\|x\|^2 \leq Bx \cdot x \leq \lambda_d(B)\,\|x\|^2 \tag{4}$$

   où $\lambda_1(B) \leq \ldots \leq \lambda_d(B)$ sont les d valeurs propres de B numérotées par ordre croissant.

   (b) Déduire de la question précédente que f vérifie (3) avec $a_f = 0$ pour tous choix de A et b. A quelle condition sur A la propriété (3) est-elle vérifiée avec $a_f > 0$? Montrer que dans le cas $a_f > 0$ la fonction f atteint son minimum sur $\mathbb{R}^d$ en exactement un point $\hat{x}$.

   (c) En considérant la matrice $B = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$, montrer que (4) n'est pas forcément vérifiée si B n'est pas symétrique.

3. Dans cette question on s'intéresse à la généralisation de l'hypothèse (2) à ce cadre :

$$\forall x, y \in \mathbb{R}^d, \qquad \|\nabla f(x) - \nabla f(y)\| \leq L\,\|x - y\|. \tag{5}$$

   (a) Soit $B \in \mathbb{R}^{d\times d}$ une matrice symétrique positive, démontrer que

$$\forall x \in \mathbb{R}^d, \qquad \|Bx\| \leq \lambda_d(B)\,\|x\|.$$

   (b) En déduire que f vérifie (5) pour une valeur de L qu'on précisera en fonction de A.

4. On suppose que (3) est vérifiée avec $a_f > 0$. Montrer que

$$\forall n, \qquad \|X_{n+1} - \hat{x}\|^2 \ \leq\ (1 - 2a_f\delta + L^2\delta^2)\|X_n - \hat{x}\|^2.$$

   En déduire une condition suffisante sur $\delta > 0$ pour assurer la convergence de $(X_n)_n$ vers $\hat{x}$.

4 Indications

1.2 La condition nécessaire et suffisante est $f''(x) \geq a_f$ pour tout réel x. On pourra penser à une formule de Taylor Lagrange pour montrer que c'est suffisant.

1.3 On peut utiliser (1) et étudier les variations de f.

1.4 Il suffit d'utiliser le Lemme de Gronwall (dont on rappelle une preuve ci-dessous).

Démonstration du lemme de Gronwall. Si $h'(t) \leq \alpha h(t)$ pour tout $t \in \mathbb{R}_+$ alors on a

$$\forall t \in \mathbb{R}_+, \qquad h'(t) - \alpha h(t) \ \leq \ 0$$

et en multipliant pas $e^{-\alpha t}$ on obtient

$$\forall t \in \mathbb{R}_+, \qquad \big(h(t)\, e^{-\alpha t}\big)' \ = \ (h'(t) - \alpha h(t))\ e^{-\alpha t} \ \leq \ 0$$

donc $t \mapsto h(t)\, e^{-\alpha t}$ est décroissante sur $\mathbb{R}_+$ et en particulier pour tout $t \geq t_0$ on $h(t)\, e^{-\alpha t} \leq h(t_0)\, e^{-\alpha t_0}$ et donc $h(t) \ \leq \ h(t_0)\, e^{\alpha(t-t_0)}$.

1.5.a Raisonner comme dans 1.4.

1.5.b Il suffit de montrer que ϕ a une limite en $+\infty$.

1.5.c Se souvenir du résultat de 1.2 pour la décroissance de $(f'(X(t))^2$, puis utiliser 1.5.b.

2.1 On utilise ici le critère de convergence de la méthode du point fixe.

2.2 Il suffit de développer le membre de gauche et d'utiliser (1) et (2), en n'oubliant pas que le minimiseur $\hat{x}$ est un point critique.

2.3.a Ne pas oublier la CNS obtenue en 1.2.

2.3.b Il faut majorer le terme $(X_{n+1} - X_n)^2$ par celui d'une série téléscopique.

2.3.c En utilisant (1) on peut obtenir $(X_{n+1} - \hat{x})^2 \leq (X_{n+1} - \hat{x})(X_n - \hat{x})$ puis conclure.

3.1 Il suffit de développer $f(x + tu)$ en un polynôme du second degré en t.

3.2.a Diagonaliser B dans une base orthonormée.

3.2.b Pour la dernière partie, ne pas oublier que f prend des valeurs positives.

3.3.a On peut utiliser 3.2.a avec la matrice ${}^t\!B\, B$.

3.4 Raisonner comme en 2.2, en utilisant plutôt (3) et (5) et le fait que $\|x\|^2 = x \cdot x$.