Méthode du gradient à pas optimal

Perrine Jouteur

Ce développement est utile pour les leçons 219, 226, 229 et 253. On peut le trouver dans le livre d'analyse numérique de Ciarlet, que je recommande De plus, un développement très proche est détaillé dans le livre *Analyse pour l'agrégation*, de Bernis et Bernis.
Attention, c'est un développement long et riche qui demande une bonne préparation.

1 Prérequis

Proposition 1.1 Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction de classe C^1 et strictement convexe sur $\mathbb{R}$. Alors f admet un unique minimum sur $\mathbb{R}$.

2 Contexte

Cette méthode s'applique pour résoudre un problème d'optimisation (on va dire de minimisation pour fixer les idées), sans contrainte c'est-à-dire qu'on veut le résoudre sur l'espace vectoriel tout entier.
En général, un problème d'optimisation sans contrainte est la donnée d'une fonction $J : V \longrightarrow \mathbb{R}$ avec V un espace vectoriel. Il s'agit de trouver $u \in V$ tel que

$$J(u) = \inf_V J(v)$$

Un cas particulier est quand V est un espace de Hilbert et J est quadratique, c'est-à-dire lorsque

$$J : u \longmapsto \frac{1}{2}a(u,u) - b(u)$$

où a est une forme bilinéaire continue symétrique et b une forme linéaire continue. Dans le cas où $V = \mathbb{R}^n$, cela se reformule par théorème de Riesz en disant que

$$J(u) = \frac{1}{2} < Au, u > - < b, u >$$

où A est une matrice symétrique et b un vecteur de $\mathbb{R}^n$. En supposant que A est définie positive, on a que J est strictement convexe.

L'existence d'une solution à un problème d'optimisation peut être obtenue grâce à de la coercivité. C'est le théorème suivant.

Théorème 2.1 Soit $J : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue et coercive. Alors il existe $u \in \mathbb{R}^n$ tel que $J(u) = \inf_V J(v)$.

En dimension infinie, c'est un peu plus compliqué. C'est d'ailleurs l'objet d'un développement, où on montre qu'une fonctionnelle continue, convexe et coercive d'un espace de Hilbert atteint son infimum.
On introduit aussi la notion de fonctionnelle elliptique.

Définition 2.1 On suppose que l'espace vectoriel V est un espace de Hilbert. Soit $J : V \longrightarrow \mathbb{R}$. On dit que J est elliptique lorsque J est de classe C^1 et qu'il existe $\alpha > 0$ telle que pour tout $u, v \in V$,

$$\langle \nabla J(v) - \nabla J(u), v - u \rangle \geqslant \alpha \left\| v - u \right\|^2$$

Par exemple, si A est symétrique définie positive, alors $u \mapsto \frac{1}{2}\langle Au, u\rangle - \langle b, u\rangle$ est elliptique.

Théorème 2.2 On suppose toujours que V est un espace de Hilbert. Soit $J : V \longrightarrow \mathbb{R}$ une fonctionnelle elliptique.
Alors J est strictement convexe et coercive, et pour tous $u, v \in V$, on a

$$J(v) - J(u) \geqslant \langle \nabla J(u), v - u \rangle + \frac{\alpha}{2} \|v - u\|^2$$

Démonstration
Faire un développement de Taylor avec reste intégral à l'ordre 0. ■

Revenons au cas de la dimension finie. On suppose donc que $V = \mathbb{R}^n$. La méthode de la relaxation consiste à prendre comme direction de descente les directions canoniques (les coordonnées) successivement (ça tourne !). Si J est elliptique, la méthode de la relaxation converge. Dans ce théorème, la dimension finie est primordiale. Et quand on applique la méthode de la relaxation pour minimiser la fonctionnelle $u \mapsto \langle Au, u \rangle - \langle b, u \rangle$, avec A symétrique définie positive, on retrouve la méthode de Gauss-Seidel pour la résolution du système $Au = b$.

La méthode de gradient à pas optimal consiste à choisir la direction comme l'opposée du gradient à chaque étape. Cela correspond à la pente la plus accentuée. Formellement, on a la définition suivante.

Définition 2.2 Soit $V = \mathbb{R}^n$. Soit $J : V \longrightarrow \mathbb{R}$ une fonctionnelle de classe C^1. La suite de la méthode du gradient à pas optimal est définie par

$$\begin{cases} u_0 \in V \\ \forall k \in \mathbb{N},\ \rho_k \in \mathbb{R} \text{ est choisi tq } J(u_k - \rho_k \nabla J(u_k)) = \inf\limits_{\rho \in \mathbb{R}} J(u_k - \rho \nabla J(u_k)) \\ u_{k+1} = u_k - \rho_k \nabla J(u_k) \end{cases}$$

3 Développement

Théorème 3.1
Soit $J : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonctionnelle elliptique.
Alors la méthode du gradient à pas optimal converge vers l'unique minimum de J, qu'on note $J(\overline{u})$.

Démonstration
Remarquons déjà que s'il existe k tel que $\nabla J(u_k) = 0$, alors la suite $(u_k)_k$ est stationnaire donc la méthode converge. On suppose donc que le gradient ne s'annule jamais au cours du processus.

• Étape 1 : la suite est bien définie.
Montrons que la suite $(u_k)_k$ est bien définie, par récurrence sur k.
Le premier terme, u_0, est bien défini.
Supposons que les termes $u_0, \cdots, u_k$ soient bien définis. La fonction

$$\phi_k : \mathbb{R} \longrightarrow \mathbb{R} \ ; \ \rho \longmapsto J(u_k - \rho \nabla J(u_k))$$

est de classe C^1 sur $\mathbb{R}$ et strictement convexe. Ainsi (prop 1) elle admet un unique minimum, atteint en $\rho_k \in \mathbb{R}$. On peut alors définir u_{k+1} sans ambiguïté, et par récurrence la suite $(u_k)_k$ est bien définie.

• Étape 2 : relation d'orthogonalité.
Le minimum ρ_k est caractérisé par $\phi_k'(\rho_k) = 0$, c'est-à-dire

$$-\Big\langle\ \nabla J(u_k - \rho_k \nabla J(u_k)), \nabla J(u_k)\ \Big\rangle = 0$$

D'où $\langle \nabla J(u_{k+1}), \nabla J(u_k) \rangle = 0$. Les directions que l'on choisit au fur et à mesure sont successivement orthogonales. cf Bernis : la droite qui passe par u_k et u_{k+1} est tangente à la ligne de niveau de J passant par u_{k+1}.

• Étape 3 : la suite $\nabla J(u_k) - \nabla J(u_{k+1})$ tend vers 0.
Par le théorème 2.2, on a

$$J(u_k) - J(u_{k+1}) \geqslant \langle \nabla J(u_{k+1}), u_k - u_{k+1} \rangle + \frac{\alpha}{2} \|u_{k+1} - u_k\|^2 = \frac{\alpha}{2} \|u_k - u_{k+1}\|^2$$

La suite $(J(u_k))_k$ est décroissante par définition, et minorée par $J(\overline{u})$, donc elle converge, donc elle est de Cauchy et en particulier $J(u_k) - J(u_{k+1}) \to 0$. D'où, en reportant dans l'inégalité ci-dessus,

$$\|u_k - u_{k+1}\| \longrightarrow 0 \quad (1)$$

Ensuite, comme J est coercive et que la suite $(J(u_k))$ est décroissante, la suite $(u_k)_k$ est nécessairement bornée. Prenons K un compact de $\mathbb{R}^n$ qui contient la suite $(u_k)_k$. Comme J est de classe C^1, sa différentielle est continue sur K donc uniformément continue sur K, et ainsi grâce à (1),

$$\|\nabla J(u_k) - \nabla J(u_{k+1})\| \longrightarrow 0$$

• Étape 4 : conclusion.
Par orthogonalité,

$$\|\nabla J(u_k)\|^2 = \langle \nabla J(u_k), \nabla J(u_k) - \nabla J(u_{k+1})\rangle \leqslant \|\nabla J(u_k)\| \, \|\nabla J(u_k) - \nabla J(u_{k+1})\|$$

d'où

$$\|\nabla J(u_k)\| \leqslant \|\nabla J(u_k) - \nabla J(u_{k+1})\| \longrightarrow 0$$

Enfin, par ellipticité de J et comme $\nabla J(\overline{u}) = 0$, on a pour tout k entier naturel,

$$\alpha \|u_k - \overline{u}\|^2 \leqslant \langle \nabla J(u_k) - \nabla J(\overline{u}), u_k - \overline{u}\rangle \leqslant \|\nabla J(u_k)\| \, \|u_k - u\|$$

Puis en simplifiant des deux côtés,

$$\|u_k - \overline{u}\| \leqslant \frac{1}{\alpha} \|\nabla J(u_k)\| \longrightarrow 0$$

On a bien la convergence voulue. ■

4 Bonus

La proposition suivante peut aussi faire l'objet d'un développement à part entière (ou pourra le trouver dans le livre de Bernis et Bernis).

Proposition 4.1
Soit A une matrice symétrique définie positive, $b \in \mathbb{R}^n$ et $J : \mathbb{R}^n \longrightarrow \mathbb{R}$ la fonctionnelle quadratique définie par A et b, i.e. $J(u) = \frac{1}{2}\langle Au, u\rangle - \langle b, u\rangle$.
Alors la méthode du gradient à pas optimal converge et pour tout entier naturel k,

$$\|x_{k+1} - \overline{x}\| \leqslant \sqrt{\frac{\lambda_n}{\lambda_1}} \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^{k+1} \|x_0 - \overline{x}\|$$

où λ_n est la plus grande valeur propre de A et λ_1 sa plus petite.