Algorithme du gradient conjugué.

• La lettre n désigne un entier supérieur ou égal à 1, A une matrice réelle carrée d'ordre n symétrique définie positive et b un vecteur de $\mathbb{R}^n$. On note $\big(\bullet\,,\,\bullet\big)$ le produit scalaire usuel dans $\mathbb{R}^n$.

1) Montrer que la résolution du système linéaire

(1) $\qquad A\,x \;=\; b$

équivaut à rechercher le minimum, pour x appartenant à $\mathbb{R}^n$, de la fonctionnelle

(2) $\qquad J(x) \;=\; \dfrac{1}{2}(x\,,\,Ax) - (b\,,\,x)\,.$

2) **Description de l'algorithme du gradient conjugué.**

• Initialisation : $\qquad x_0 \in \mathbb{R}^n\,, \quad w_0 \;=\; Ax_0 - b\,.$

• Itération de l'algorithme.

On suppose connus l'état x^{k-1} après la $k^{\text{ième}}$ itération ainsi que la direction de descente w^{k-1}. L'état x^k au pas suivant de l'algorithme est issu de x^{k-1} via un incrément proportionnel à la direction de descente :

(3) $\qquad x^k \;=\; x^{k-1} \,+\, \rho^{k-1}\, w^{k-1}$

de façon à minimiser la fonctionnelle J :

(4) $\qquad J(x^k) \;\leq\; J\big(x^{k-1} \,+\, \rho\, w^{k-1}\big)\,, \quad \forall \rho \in \mathbb{R}\,.$

On introduit également le gradient g^k de J au point x^k :

(5) $\qquad g^k \;=\; Ax^k \,-\, b\,.$

2.1) Calculer le coefficient ρ^{k-1} en fonction de g^{k-1}, w^{k-1} et A.

2.2) Montrer que l'on a

(6) $\qquad \big(g^k\,,\,w^{k-1}\big) \;=\; 0\,.$

La nouvelle direction de descente w^k est recherchée sous la forme

(7) $\qquad w^k \;=\; g^k \,+\, \alpha^k w^{k-1}$

de sorte que w^k soit orthogonal à la direction w^{k-1} pour le produit scalaire associé à la matrice A c'est à dire :

(8) $\qquad \big(w^k\,,\,Aw^{k-1}\big) \;=\; 0\,.$

2.3) Calculer la valeur de α^k en fonction de g^k, w^{k-1} et A. On notera que le choix plus simple $\alpha^k = 0$ conduit à la méthode du gradient simple, où la direction de descente de l'algorithme w^k correspond à la ligne de plus grande pente g^k.

3) L'algorithme converge en au plus n étapes.

3.1) Remarquer que si les gradients successifs $g^0, g^1, \ldots, g^{k-1}$ sont non nuls mais que g^k est nul, alors l'état x^k est la solution du système linéaire (1).

3.2) On suppose dans cette question que tous les gradients g^j sont non nuls jusqu'à l'étape m incluse. Montrer qu'on a alors les relations d'orthogonalité suivantes :

(9) $\quad \big(g^k, w^j\big) = 0, \quad 0 \leq j < k \leq m$

(10) $\quad \rho^k \neq 0, \quad 0 \leq k \leq m$

(11) $\quad \big(g^k, g^j\big) = 0, \quad 0 \leq j < k \leq m$

(12) $\quad \big(w^k, Aw^j\big) = 0, \quad 0 \leq j < k \leq m.$

On pourra raisonner par récurrence sur k pour l'ensemble des quatre propriétés et démontrer les relations (9) à (12) dans l'ordre où elles apparaissent.

3.3) Montrer que l'algorithme converge en au plus n itérations.

4) Propriétés auxiliaires.

4.1) Montrer que l'on a

(13) $\quad \alpha^k = \dfrac{\| g^k \|^2}{\| g^{k-1} \|^2}\,.$

En déduire que le calcul très précis des produits scalaires est crucial pour assurer le succès pratique (informatique) de la convergence de la méthode.

4.2) Montrer que l'état x^k réalise le minimum de la fonctionnelle J sur le sous-espace affine $x^0 + < w^0, \ldots, w^{k-1} >$ passant par le point x^0 et d'espace vectoriel associé engendré par les vecteurs $w^0, \ldots, w^{k-1}$. Quel commentaire pouvez-vous faire ?

FD, octobre 1993, janvier 1995, août 1999, juillet 2002, novembre 2006.

Algorithme du gradient conjugué.
Proposition de corrigé.

1) La fonctionnelle J a bien un minimum car la matrice A est symétrique définie positive. L'équation d'Euler au point de minimum s'écrit ici :
$J'(x) \bullet (Ax - b) \,=\, 0\,,$ ce qui montre la propriété.

2.1) On a simplement

$$(14) \qquad \rho^{k-1} \;=\; -\,\frac{(g^{k-1}\,,\,w^{k-1})}{(w^{k-1}\,,\,Aw^{k-1})}\,.$$

2.2) La condition précédente a été obtenue en écrivant qu'au point x^k, la fonctionnelle J est minimale si on la restreint à la droite affine passant par x^{k-1} et dirigée par le vecteur w^{k-1} ce qui s'exprime en annulant le gradient de J au point x^k dans la direction w^{k-1}, ce qui constitue exactement la condition (6).

2.2) On a immédiatement

$$(15) \qquad \alpha^{k} \;=\; -\,\frac{(g^{k}\,,\,Aw^{k-1})}{(w^{k-1}\,,\,Aw^{k-1})}\,.$$

3.1) Dès que l'un des gradients g^m est nul, on est en un point x^m qui est solution du système (1) et l'algorithme ne présente plus alors aucun intérêt puisque l'on a résolu le problème posé.

3.2) On montre la propriété par récurrence.

- La propriété est vraie pour $k = 1$. On remarque que $w^0 \,=\, g^0$ est non nul en vertu de la question précédente. On a ensuite $(g^1\,,\,w^0) \,=\, 0$ compte tenu de la relation (6). Si on considère la relation (7) pour $k = 1$ et qu'on la multiplie scalairement par g^1, on obtient $(g^1\,,\,w^1) = \parallel g^1 \parallel^2$, ce qui montre que w^1 est non nul puisque g^1 est non nul. Le numérateur comme le dénominateur de l'expression (14) qui permet de calculer ρ^0 sont non nuls, donc ρ^0 est non nul et la relation (10) est vraie à l'ordre $k \,=\, 0$. La relation (11) est une conséquence simple du choix de la direction de descente initiale : $(g^1\,,\,g^0) \,=\, (g^1\,,\,w^0) \,=\, 0$ compte tenu de la relation (9). Enfin, la relation (12) exprime simplement la relation (8) pour $k = 1$.
- On suppose les relations (9) à (12) vérifiées jusqu'à l'ordre k inclus et on les étend pour l'indice $k+1$, en supposant g^{k+1} non nul. On remarque d'abord que la relation (3) entraîne clairement

$$(16) \qquad g^{k+1} \;=\; g^k \,+\, \rho^k\, Aw^k\,.$$

On a d'une part $(g^{k+1}\,,\,w^k) \,=\, 0$ compte tenu de la relation (6) et pour $j < k$, on a d'autre part $(g^{k+1}\,,\,w^j) \,=\, (g^{k+1} - g^k\,,\,w^j) + (g^k\,,\,w^j) \,=\, \rho^k\,(Aw^k\,,\,w^j)$ compte tenu de (16) et de l'hypothèse de récurrence (9). Or cette dernière expression est nulle en vertu de l'hypothèse de récurrence (12), donc la relation (9) est établie. Si on multiplie scalairement l'identité (7) écrite au rang $k+1$ par g^{k+1}, la relation (9) que nous venons de démontrer entraîne que l'on a

(17) $\qquad (g^{k+1}\,,\, w^{k+1}) \;=\; \| \, g^k \, \|^2 \,,$

et la relation (10) est alors une conséquence simple de la relation qui permet de calculer ρ^k. Exprimons g^j grâce à la relation (7) (considérée avec $k = j$). Il vient $(g^{k+1}\,,\, g^j) \;=\; (g^{k+1}\,,\, w^j) \,-\, \alpha^j\,(g^{k+1}\,,\, w^{j-1})$ et cette expression est nulle compte tenu de la relation (9) prise à l'ordre $k+1$. Ceci montre la relation (11). La relation (12) est la conséquence de la relation (8) pour $j = k$ et du calcul suivant pour $j < k$:

$$\begin{aligned} (w^{k+1}\,,\, Aw^j) \;&=\; (g^{k+1}\,,\, Aw^j) && \text{compte tenu de (7) et de (12)} \\ &=\; \Big(g^{k+1}\,,\, \frac{1}{\rho^j}\,(g^{j+1} - g^j)\,\Big) && \text{en vertu de (16)} \\ &=\; 0 && \text{compte tenu de (11) .} \end{aligned}$$

La propriété est donc démontrée par récurrence. $\square$

3.3) Compte tenu de la relation (11), la suite de gradients $g^0\,,\, g^1\,,\, \ldots\,,\, g^k$ est composée de vecteurs orthogonaux deux à deux jusqu'à un ordre m et ceci a lieu dans l'espace $\mathbb{R}^n$. La conclusion est alors claire.

4.1) Dans la relation (15), on exprime le vecteur Aw^{k-1} au numérateur et au dénominateur à l'aide de la relation (16) et on utilise la relation (11) pour le numérateur. Il vient :

$$\alpha^k \;=\; \frac{\| \, g^k \, \|^2}{(g^{k-1}\,,\, w^{k-1})}$$

et la relation (13) est alors conséquence directe de la relation (17).

4.2) Si on écrit l'inéquation d'Euler qui caractérise le minimum de la fonction convexe dérivable J sur le convexe fermé $x^0 + \, < w^0\,,\, \ldots\, w^{k-1} >$, atteint au point $y \in \mathbb{R}^n$ il vient $J'(y) \bullet w^j \;=\; 0$ pour tout $j = 1, 2, ..., k-1$. Mais le choix $y = x^k$ vérifie ces relations puisque d'une part x^k appartient au convexe $x^0 + \, < w^0\,,\, \ldots\, w^{k-1} >$, compte tenu de l'initialisation de l'algorithme et de la relation constitutive (3) et d'autre part $J'(x^k) \bullet w^j \;=\; (g^k\,,\, w^j)$ lequel est nul en vertu de la relation (9). $\square$

• Cette propriété montre que le point x^k, conçu initialement à la relation (4) pour minimiser la fonctionnelle J le long de la droite affine passant par x^{k-1} et dirigée selon le vecteur w^{k-1} la minimise en fait sur tout un sous espace de dimension k ! On touche là au génie de Hestenes et Stiefel qui ont inventé la méthode du gradient conjugué en 1952.

FD, octobre 1993, janvier 1995, août 1999, juillet 2002, novembre 2006.