Méthode du gradient conjugué

Notations
Soit A une matrice symétrique définie positive sur $\mathbb{R}^n$, b un vecteur de $\mathbb{R}^n$. On note $\bar{x} = A^{-1}b$ et $x_0 \in \mathbb{R}^n$, et on définie la fonctionnelle

$$J(x) = \frac{1}{2}(Ax, x) - (b, x)$$

Dans la suite, si on dispose d'une suite $(x_k)_k$ de $\mathbb{R}^n$, on note $G_k = \mathrm{Vect}(\nabla J(x_i), i = 0..k)$. On notera $c = \lambda_n(A)/\lambda_1(A) = \mathrm{Cond}_2(A)$ le conditionnement de A.

Théorème. *A x_0 fixé, il existe une unique suite $(x_k)_{k\geq 0}$ telle que pour tout k,*

$$x_{k+1} = \mathit{argmin}(J_{|x_0+G_k})$$

Celle-ci vérifie la relation de récurrence suivante :

$$\begin{cases} d_k = \nabla J(x_k) + \frac{\|\nabla J(x_{k-1})\|^2}{\|\nabla J(x_k)\|^2} d_{k-1} \qquad \textit{(avec la convention que } \; d_{-1} = 0) \\ r_k = \frac{(\nabla J(x_k), d_k)}{(Ad_k, d_k)} \\ x_{k+1} = x_k - r_k d_k \end{cases}$$

Et cette suite stationne sur $\bar{x}$ en moins de n étapes. Enfin, on a pour tout k :

$$G_k = \mathbb{R}_{k-1}[A]\nabla J(x_0) = A\mathbb{R}_{k-1}[A](x_0 - \bar{x})$$

Démonstration. J est une fonction continue, strictement convexe et coercive sur $\mathbb{R}^n$, donc sur tout les sous-espaces affines de $\mathbb{R}^n$. Ainsi, si $x_0, ..x_k$ sont construit de manière unique, alors x_{k+1} l'est aussi par l'étude de $J_{|x_0+G_k}$. De plus, x_{k+1} est atteint en l'unique point critique de $J_{|x_0+G_k}$, c'est-à-dire que

$$0 = \nabla(J_{|x_0+G_k})(x_{k+1}) = \mathrm{proj}^{\perp}_{G_k}(\nabla J(x_{k+1}))$$

x_{k+1} est donc définit par les $k+1$ relations :

$$(\nabla J(x_i), \nabla J(x_{k+1})), \quad i = 0..k \tag{1}$$

On suppose dans la suite que la suite n'est pas encore stationnaire, c'est-à dire que les $x_0, ..x_{k+1}$ sont distincts. On note $\Delta_l = x_{l+1} - x_l \in G_l$. On a $\nabla J(x_{l+k}) = \nabla J(x_k) + A\Delta_k$; en injectant cela dans (1) :

$$(\Delta_l, A\Delta_k), \quad i = 0..k \tag{2}$$

Les $(\Delta_l)_l$ forment donc une famille orthogonale pour le produit scalaire $(\cdot, A\cdot)$. C'est donc une famille libre et par un argument de dimension, on a, pour tout l,

$$\mathrm{Vect}(\nabla J(x_i), i = 0..l) = \mathrm{Vect}(\Delta_i = 0..l)$$

En particulier, $(\Delta_k, \nabla J(x_k)) \neq 0$ car sinon $\mathrm{Vect}(\Delta_i = 0..k) \subset \mathrm{Vect}(\nabla J(x_i), i = 0..k-1)$. Cela permet d'écrire $x_{k+1} = x_k - r_k d_k$ (donc $\Delta_k = -r_k d_k$) où $r_k \in \mathbb{R}$ est choisi tel que la décomposition de d_k dans la base $\nabla J(x_0), ..\nabla J(x_k)$ est de la forme :

$$d_k = \nabla J(x_k) + \sum_{l=0}^{k-1} \lambda_l^k \nabla J(x_l)$$

Il s'agit alors de calculer les λ_l^k. Soit l dans $0..k$, on a :

$$\begin{aligned}
0 &= (Ad_k, \Delta_l) \\
&= (d_k, A\Delta_l) \\
&= (\nabla J(x_k) + \sum_{l=0}^{k-1} \lambda_l^k \nabla J(x_l), \nabla J(x_{l+1}) - \nabla J(x_l)) \\
&= \begin{cases} \lambda_{l+1}^k \|\nabla J(x_{l+1})\|^2 - \lambda_l^k \|\nabla J(x_l)\|^2 & \text{si l<k-1} \\ \|\nabla J(x_k)\|^2 - \lambda_k^k \|\nabla J(x_{k-1})\|^2 & \text{si l=k-1} \end{cases}
\end{aligned}$$

De proche en proche, on a donc $\lambda_l^k = \frac{\|\nabla J(x_k)\|^2}{\|\nabla J(x_l)\|^2}$. Cela fourni une relation de récurrence miraculeuse pour d_k ; en effet

$$d_k = \nabla J(x_k) + \sum_{l=0}^{k-1} \frac{\|\nabla J(x_k)\|^2}{\|\nabla J(x_l)\|^2} \nabla J(x_l) = \nabla J(x_k) + \frac{\|\nabla J(x_k)\|^2}{\|\nabla J(x_{k-1})\|^2} d_{k-1}$$

Il reste à calculer r_k.
$J(x_k - rd_k) = \frac{1}{2}(d_k, Ad_k)r^2 - (\nabla J(x_k), d_k)r + J(x_k)$ est une parabole qui atteint son minimum en $r_k = \frac{(\nabla J(x_k), d_k)}{(d_k, Ad_k)}$.
Enfin la propriété que $G_k = \mathbb{R}_{k-1}[A]\nabla J(x_0) = A\mathbb{R}_{k-1}[A](x_0 - \bar{x})$ à tout rang vient de la décomposition :

$$\begin{aligned}
\nabla J(x_{k+1}) &= A(x_k - r_k d_k) \\
&= A(x_k - x_0) + A(x_0 - \bar{x}) - r_k A d_k \\
&\in AG_{k-1} + A(x_0 - \bar{x}) + AG_k
\end{aligned}$$

□

Référence : Ciarlet