

31 Méthode du gradient conjugué

Ref : Allaire, Dumas, Ciarlet.

THÉORÈME 31.1 Soit $A \in S^{++}(n, \mathbb{R})$ et $b \in \mathbb{R}^n$. On prend $x_0 \in \mathbb{R}^n$ quelconque et on définit $r_0 = b - Ax_0$ et ses sous-espaces vectoriels (dits de Krylov) associés :

$$K_m(r_0) = \text{Vect}\{r_0, \dots, A^m r_0\}$$

Il existe une unique suite (x_m) définie par récurrence par les conditions :

$$x_m \in x_0 + K_{m-1} \quad , \quad r_m = b - Ax_m \perp K_{m-1} \quad \text{et} \quad r_m \in K_m$$

et cette suite stationne vers l'unique solution de $Ax = b$ après au plus n itérations.

idée : On introduit la fonctionnelle quadratique $J(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle$, comme A est définie positive, les niveaux de J sont l'image par une transformation affine de cercles concentriques. Il existe donc un unique minimum que l'on cherche numériquement par une méthode de descente. Ce minimum est alors un point critique de J , c'est-à-dire une solution de $\nabla J(x) = Ax - b = 0$.

PREUVE. Remarquons tout d'abord que la suite des sous-espaces vectoriels (K_m) est strictement croissante pour l'inclusion puis stationne. En effet, la suite est croissante et doit stationner car $\mathbb{R}^n$ est de dimension finie. De plus, si i désigne le plus petit indice tel que $K_{m_0} = K_{m_0+1}$ alors K_{m_0} est stable par A puisque $A(A^{m_0} r_0) \in K_{m_0}$ et (K_m) stationne alors à partir de m_0 .

On cherche x_m sous la forme $x_m = x_0 + y_m$ avec $y_m \in K_{m-1}$. Remarquons que $r_m = Ax_0 + Ay_m - b = r_0 + Ay_m \in K_m$.

La condition d'orthogonalité s'écrit :

$$\begin{aligned} \langle b - Ax_m, y \rangle &= 0, \forall y \in K_m \Leftrightarrow \langle b - Ax_0 - Ay_m, y \rangle = 0, \forall y \in K_m \\ &\Leftrightarrow \langle A^{-1}r_0 - y_m, y \rangle_A = 0, \forall y \in K_m \end{aligned}$$

C'est-à-dire, y_m est la projection orthogonale de $A^{-1}r_0$ sur K_m pour le produit scalaire défini par A , qui est bien défini de manière unique. D'où l'existence et l'unicité de x_m .

Au rang critique k_0 , on trouve $r_{m_0+1} \in K_{m_0+1} = K_{m_0}$ mais aussi $r_{m_0+1} \perp K_{m_0}$, c'est-à-dire, $r_{m_0+1} = 0$ et x_{m_0+1} est solution du système linéaire et $m_0 + 1 \leq n$, on a fait au plus n itérations. $\square$

Cette preuve est théorique et ne donne pas immédiatement d'algorithme pour mettre la méthode en pratique. Voyons comment le mettre en oeuvre.

Au cours de la preuve, on a vu que (r_m) est une famille orthogonale pour le produit scalaire usuel et que $K_m = \text{Vect}(r_0, \dots, r_m)$. Introduisons les directions de descentes $d_m = x_{m+1} - x_m \in K_m$. La famille d_k est A -orthogonale : en effet une récurrence immédiate donne : pour $m \geq l$,

$$\langle Ad_{m+1}, d_l \rangle = \langle Ax_{m+2} - Ax_{m+1}, d_l \rangle = \langle r_{m+2} - r_{m+1}, d_l \rangle = 0$$

La famille (d_m) apparait alors comme une orthogonalisée de Gram-Schmidt des (r_m) pour le produit scalaire défini par A .

Expliquer sur un dessin, la différence entre la direction du gradient et la direction conjuguée : en dimension 2, on termine en 2 coups. Voici donc l'algorithme :

Soit $p_0 = r_0$ et pour $m \geq 0$,

$$p_m = r_m - \sum_{j=0}^{m-1} \frac{\langle r_m, p_j \rangle_A}{\langle p_j, p_j \rangle_A} p_j = r_m - \sum_{j=0}^{m-1} \beta_j p_j$$

où

$$\beta_j = \frac{\langle r_m, p_j \rangle_A}{\langle p_j, p_j \rangle_A}$$

C'est une famille orthogonalisée de Gram-Schmidt de (r_m) , elle diffère de d_m par des constantes positives : $\exists \alpha_m > 0$ tel que :

$$d_m = x_{m+1} - x_m = \alpha_m p_m$$

On cherche alors une méthode itérative pour calculer les β_j et les α_m .

On a :

$$r_{m+1} - r_m = A(x_m - x_{m+1}) = -\alpha_m A p_m$$

Donc :

$$\langle r_m, p_j \rangle_A = \langle r_m, \frac{r_j - r_{j+1}}{\alpha_j} \rangle$$

Puis,

$$\beta_j = 0 \text{ si } j \leq m-2, \quad = -\frac{\langle r_m, r_m \rangle}{\alpha_{m-1} \langle p_{m-1}, p_{m-1} \rangle_A} \text{ si } j = m-1$$

Enfin, α_m s'obtient par orthogonalité des r_i :

$$\langle r_{m+1} - r_m \rangle = 0 = \langle r_m - \alpha_m A p_m, r_m \rangle = \|r_m\|^2 - \alpha_m \langle A p_m, p_m + \beta_{m-1} p_{m-1} \rangle = \|r_m\|^2 - \alpha_m \|p_m\|_A^2$$

Donc

$$\alpha_m = \frac{\|r_m\|^2}{\|p_m\|_A^2}$$

On en déduit un algorithme itératif : On choisit x_0 quelconque, on pose $p_0 = b - A x_0$, puis :

$$\begin{aligned} r_{m+1} &= r_m - \frac{\|r_m\|^2}{\|p_m\|_A^2} A p_m \\ p_{m+1} &= r_{m+1} + \frac{\|r_{m+1}\|^2}{\|r_m\|_A^2} p_m \\ x_{m+1} &= x_m + \frac{\|r_{m+1}\|^2}{\|p_{m+1}\|_A^2} p_m \end{aligned}$$

Remarque : L'indice m est plus pertinent que k pour éviter de prononcer "caca" qui peut déstabiliser le jury et soi-même.

Leçons concernées : Problème d'extremum, $f(x)=0$, systèmes linéaires.