

JEAN HARDOUIN DUPARC

**Méthode de résolution de système linéaire
cramérien par projections successives avec
orthogonalisation des restes**

Revue française d'informatique et de recherche opérationnelle. Série rouge, tome 3, n° R2 (1969), p. 15-25

http://www.numdam.org/item?id=M2AN_1969__3_2_15_0

© AFCET, 1969, tous droits réservés.

L'accès aux archives de la revue « Revue française d'informatique et de recherche opérationnelle. Série rouge » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

METHODE DE RESOLUTION DE SYSTEME LINEAIRE CRAMERIEN PAR PROJECTIONS SUCCESSIVES AVEC ORTHOGONALISATION DES RESTES

par Jean HARDOUIN DUPARC

Résumé. — La méthode décrite dans cet article permet de résoudre un système d'équations linéaires cramérien en effectuant un nombre d'itérations au plus égal au nombre d'équations, quelle que soit la matrice des coefficients. Cette méthode paraît avantageuse dans le cas où la matrice des coefficients comporte un petit nombre d'éléments non nuls et où la matrice obtenue en multipliant par la transposée, n'a pas cette qualité.

Cette méthode s'inspire de la méthode de décomposition de la norme euclidienne mais au lieu de prendre comme vecteur projetant un vecteur de la forme : ${}^T A \cdot Z$ avec $Z = R / \|R\|$ (où A désigne la matrice des coefficients, Z un vecteur et R le vecteur Reste), nous prenons une forme analogue mais où le vecteur Z évolue de telle sorte que les R soient orthogonaux deux à deux.

Ainsi, cette méthode serait à la méthode de décomposition de la norme euclidienne, ce qu'est la méthode du gradient conjugué à celle de la plus profonde descente.

Nous démontrerons d'abord la convergence de la méthode en n itérations au plus, puis nous comparerons à la méthode du gradient conjugué appliquée au système transformé par multiplication préalable de la matrice des coefficients et du second membre par la matrice transposée.

Cette méthode paraît bien adaptée à la résolution de systèmes linéaires de dimension moyenne, particulièrement dans le cas où la matrice des coefficients est lacunaire sans que d'autres méthodes plus rapides soient applicables, telle, par exemple, la méthode de surelaxation.

EXPOSE DE LA METHODE

Soit à résoudre l'équation $AX = B$ avec :

$$\left. \begin{array}{l} A : \text{Matrice carrée } (n, n) \text{ non singulière, donnée} \\ B : \text{Vecteur donné } (n) \\ X : \text{Vecteur inconnu } (n) \end{array} \right\} \begin{array}{l} A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n) \\ \det(A) \neq 0 \\ B, X \in \mathbb{R}^n \end{array}$$

Nous considérons les vecteurs suivants :

$$X_i, R_i, V_i, S_i, \in \mathbb{R}^n$$

i étant un indice variant à partir de 0.

Au départ, nous initialisons :

X_0 à un vecteur quelconque (par exemple le vecteur nul)

$$R_0 = A \cdot X_0 - B$$

$$V_0 = R_0$$

$$S_0 = {}^T A \cdot V_0$$

Nous répétons ensuite, tant que $\|S_i\| \neq 0$ la suite de calculs suivants :

$$X_{i+1} = X_i - \frac{{}^T R_i \cdot V_i}{\|S_i\|^2} S_i$$

$$R_{i+1} = R_i - \frac{{}^T R_i \cdot V_i}{\|S_i\|^2} A \cdot S_i \quad (1)$$

$$V_{i+1} = R_{i+1} - \frac{{}^T S_i \cdot {}^T A R_{i+1}}{\|S_i\|^2} V_i \quad (2)$$

$$S_{i+1} = {}^T A \cdot R_{i+1} - \frac{{}^T S_i \cdot {}^T A R_{i+1}}{\|S_i\|^2} S_i$$

On voit que : $S_i = {}^T A \cdot V_i$ entraîne $S_{i+1} = {}^T A \cdot V_{i+1}$, cette relation est donc générale puisque $S_0 = {}^T A V_0$

Démonstration

Nous pouvons récrire les formules (1) et (2) en ne conservant que les vecteurs R_i et V_i :

$$R_{i+1} = R_i - \frac{{}^T R_i \cdot V_i}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} A \cdot {}^T A \cdot V_i \quad (3)$$

$$V_{i+1} = R_{i+1} - \frac{{}^T V_i \cdot A \cdot {}^T A \cdot R_{i+1}}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} V_i \quad (4)$$

Il suffit de montrer que quel que soit A non singulière et $R_0 = V_0$

on a : ${}^T R_i \cdot R_j = \delta_{ij} {}^T R_i \cdot R_i$ (δ_{ij} : symbole de Kronecker)

Cette démonstration est la même que celle du théorème de Steifel-Hestenes.

Au préalable, il est nécessaire de montrer :

$$\|S_i\| = 0 \Leftrightarrow \|V_i\| = 0 \Leftrightarrow \|R_i\| = 0$$

La première relation est une conséquence du fait que A est non singulière; pour la seconde, nous savons que $R_0 = V_0$, par la suite, si nous supposons

$$\|R_i\| \neq 0 \quad \|V_i\| \neq 0 \quad \text{et} \quad \|R_{i+1}\| = 0$$

Alors A et ${}^T A$ étant non singulières, les deux éléments composants V_{i+1} dans (4) sont nuls et $\|V_{i+1}\| = 0$ et si $\|R_j\| = 0$ et $\|V_j\| = 0$ d'après (3) et (4) $\|R_{j+1}\|$ et $\|V_{j+1}\|$ sont nuls.

Ce qui justifie le test $\|S_i\| \neq 0$ dans l'algorithme.

Ainsi, l'algorithme peut être poursuivi tant que $\|S_i\| \neq 0$ et lorsque cette éventualité ne se produit plus, alors $\|R_i\| = 0$ et X_i est solution du problème.

En multipliant (3) à gauche par ${}^T V_i$, nous obtenons :

$${}^T V_i \cdot R_{i+1} = {}^T V_i \cdot R_i - \frac{{}^T R_i \cdot V_i}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} {}^T V_i \cdot A \cdot {}^T A \cdot V_i$$

Comme ${}^T V_i \cdot A \cdot {}^T A \cdot V_i$ n'est pas nul et que le produit scalaire est commutatif nous en déduisons :

quel que soit i :

$${}^T V_i \cdot R_{i+1} = 0 \tag{5}$$

ou

$${}^T R_{i+1} \cdot V_i = 0 \tag{5 bis}$$

En multipliant (4) à gauche par ${}^T R_{i+1}$, nous obtenons :

$${}^T R_{i+1} \cdot V_{i+1} = {}^T R_{i+1} \cdot R_{i+1} - \frac{{}^T V_i \cdot A \cdot {}^T A \cdot R_{i+1}}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} {}^T R_{i+1} \cdot V_i$$

D'où en vertu de (5 bis) :

quel que soit i :

$${}^T V_i \cdot R_i = {}^T R_i \cdot V_i = \|R_i\|^2 \tag{6}$$

En multipliant (4) à gauche par ${}^T R_i$, nous obtenons :

$${}^T R_i \cdot V_{i+1} = {}^T R_i R_{i+1} - \frac{{}^T V_i \cdot A \cdot {}^T A \cdot R_{i+1}}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} {}^T R_i \cdot V_i$$

ou en appliquant la commutativité du produit scalaire :

$${}^T R_i V_{i+1} = {}^T R_{i+1} R_i - \frac{{}^T R_i \cdot V_i}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} {}^T R_{i+1} \cdot A \cdot {}^T A \cdot V_i$$

ou en mettant ${}^T R_{i+1}$ en facteur :

$${}^T R_i V_{i+1} = {}^T R_{i+1} \left\{ R_i - \frac{{}^T R_i \cdot V_i}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} A \cdot {}^T A \cdot V_i \right\}$$

D'où en remplaçant la quantité entre accolades par sa valeur en vertu de (3), nous aurons :
quel que soit i

$${}^T R_i \cdot V_{i+1} = \|R_{i+1}\|^2 \quad (7)$$

ou

$${}^T V_{i+1} \cdot R_i = \|R_{i+1}\|^2 \quad (7 \text{ bis})$$

et en rapprochant (6), (7) et (7 bis)

quel que soit i

$${}^T V_{i+1} \cdot R_{i+1} = {}^T V_{i+1} \cdot R_i = \|R_{i+1}\|^2 \quad (8)$$

$${}^T R_{i+1} \cdot V_{i+1} = {}^T R_i \cdot V_{i+1} = \|R_{i+1}\|^2 \quad (8 \text{ bis})$$

En multipliant (3) à gauche par ${}^T V_{i+1}$, nous obtenons :

$${}^T V_{i+1} \cdot R_{i+1} = {}^T V_{i+1} \cdot R_i - \frac{{}^T R_i \cdot V_i}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} {}^T V_{i+1} \cdot A \cdot {}^T A \cdot V_i$$

Et en appliquant (8) :
quel que soit i .

$${}^T V_{i+1} \cdot A \cdot {}^T A \cdot V_i = 0 \quad (9)$$

Pour la commodité, nous réécrivons maintenant les formules (3) et (4) sous les formes suivantes :

$$R_{i+1} = R_i - \lambda_i A \cdot {}^T A \cdot V_i \quad (10)$$

$$V_{i+1} = R_{i+1} - \sigma_i V_i \quad (11)$$

Nous allons maintenant établir par une double récurrence les formules suivantes :

quel que soit h et quel que soit $j < h$

$${}^T R_h \cdot V_j = 0 \quad (12)$$

$${}^T R_h \cdot R_j = 0 \quad (13)$$

$${}^T V_h \cdot A \cdot {}^T A \cdot V_j = 0 \quad (14)$$

ces formules sont vraies pour $h = 1$ et $j = 0$.

En effet,

$${}^T R_1 \cdot V_0 = 0 \text{ d'après (5 bis)} \quad (12 \text{ bis})$$

et $R_0 = V_0$ entraîne

$${}^T R_1 \cdot R_0 = 0 \quad (13 \text{ bis})$$

$${}^T V_1 \cdot A \cdot {}^T A V_0 = 0 \quad \text{d'après (9)} \quad (14 \text{ bis})$$

Supposons-les vraies pour $h = p$ et $j < p$.

Nous allons montrer qu'elles le sont encore pour $h = p + 1$ et $j < p + 1$ ce qui revient à montrer qu'elles sont vraies pour les couples

$$(p + 1, 0), (p + 1, 1), \dots, (p + 1, p)$$

$${}^T R_{p+1} \cdot V_0 = {}^T R_{p+1} \cdot R_0 \quad \text{puisque} \quad V_0 = R_0$$

Or, d'après (10) nous pouvons écrire

$${}^T R_{p+1} \cdot V_0 = {}^T R_p V_0 - \lambda_p {}^T V_0 \cdot A \cdot {}^T A \cdot V_p$$

Ce qui entraîne

$${}^T R_{p+1} \cdot R_0 = {}^T R_{p+1} \cdot V_0 = 0 \quad (12 \text{ ter et } 13 \text{ ter})$$

Pour q variant de 1 à p , nous pouvons établir :

en vertu de (10), nous aurons :

$$\begin{aligned} {}^T R_{p+1} \cdot V_q &= {}^T R_p \cdot V_q - \lambda_p {}^T V_p \cdot A \cdot {}^T A \cdot V_q \\ &\Rightarrow {}^T R_{p+1} \cdot V_q = 0 \end{aligned} \quad (12 \text{ quater})$$

en vertu de (11), nous aurons :

$$\begin{aligned} {}^T R_{p+1} \cdot V_q &= {}^T R_{p+1} \cdot R_q - \sigma_{q-1} {}^T R_{p+1} \cdot V_{q-1} \\ &\Rightarrow {}^T R_{p+1} \cdot R_q = 0 \end{aligned} \quad (13 \text{ quater})$$

en vertu de (10), nous aurons :

$$\begin{aligned} {}^T R_{p+1} R_q &= {}^T R_{p+1} R_{q-1} - \lambda_{q-1} {}^T R_{p+1} \cdot A \cdot {}^T A \cdot V_{q-1} \\ &\Rightarrow {}^T R_{p+1} \cdot A \cdot {}^T A \cdot V_{q-1} = 0 \end{aligned} \quad (15)$$

Puis pour q variant de 1 à $p - 1$, nous établissons par application de (11) et de (15)

$$\begin{aligned} {}^T R_{p+1} \cdot A \cdot {}^T A \cdot V_q &= {}^T V_{p+1} \cdot A \cdot {}^T A \cdot V_q - \sigma_p {}^T V_p \cdot A \cdot {}^T A \cdot V_q \\ &\Rightarrow {}^T V_{p+1} \cdot A \cdot {}^T A \cdot V_q = 0 \end{aligned} \quad (14 \text{ ter})$$

Enfin nous avons :

$${}^T V_{p+1} \cdot A \cdot {}^T A \cdot V_p = 0 \quad \text{d'après (9)} \quad (14 \text{ quater})$$

Nous voyons ainsi que si $\|S_i\| \neq 0$ pour tout $i < k$

$${}^T R_p \cdot R_q = \delta_{pq} {}^T R_p R_p \quad \text{pour tout } p \text{ et } q \leq k$$

ce qui entraîne qu'il ne peut pas y avoir plus de n vecteurs restes non nuls et qu'il existe une valeur j avec $j \leq n$ telle que l'on a $\|R_j\| = \|V_j\| = \|S_j\| = 0$ et par conséquent X_j est la solution.

SIMPLIFICATION DE L'ALGORITHME

En vertu de (8), nous pouvons réécrire (3) sous la forme

$$R_{i+1} = R_i - \frac{{}^T R_i \cdot R_i}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} A \cdot {}^T A \cdot V_i$$

En multipliant cette relation à gauche par ${}^T R_{i+1}$, nous avons

$${}^T R_{i+1} \cdot R_{i+1} = {}^T R_{i+1} \cdot R_i - \frac{{}^T R_i \cdot R_i}{{}^T V_i \cdot A \cdot {}^T A \cdot V_i} {}^T R_{i+1} \cdot A \cdot {}^T A \cdot V_i$$

Ce qui permet puisque ${}^T R_{i+1} \cdot R_i = 0$ d'écrire (4) sous la forme :

$$V_{i+1} = R_{i+1} + \frac{\|R_{i+1}\|^2}{\|R_i\|^2} V_i$$

D'où nous pouvons reprendre l'algorithme sous la forme suivante :

$$\begin{aligned} i &= 0 \\ X_i &= 0 \\ R_i &= -B \\ S_i &= {}^T A \cdot R_i \end{aligned}$$

pour $i = i + 1$ tant que $\|S_{i-1}\| \neq 0$

$$\left\{ \begin{aligned} X_i &= X_{i-1} - \frac{\|R_{i-1}\|^2}{\|S_{i-1}\|^2} S_{i-1} \\ R_i &= R_{i-1} - \frac{\|R_{i-1}\|^2}{\|S_{i-1}\|^2} A \cdot S_{i-1} \\ S_i &= {}^T A R_i + \frac{\|R_i\|^2}{\|R_{i-1}\|^2} S_{i-1} \end{aligned} \right.$$

Nous remarquerons que les vecteurs $(A \cdot S_{i-1})$ et $({}^T A \cdot R_i)$ n'ont pas besoin d'être mémorisés et la méthode est donc à trois vecteurs, par conséquent moins encombrante que la méthode du gradient conjugué.

Nous pouvons maintenant décrire l'algorithme d'exécution et écrire le programme en Algol.

Dans l'algorithme d'exécution nous entrons dans la boucle par l'étage de modification de S.

Dans le programme Algol, nous avons supposé la matrice explicitée mais il est bien évident que dans la pratique ce cas se présentera rarement.

ALGORITHME D'EXECUTION

$X := 0$
 $R := -B$
 $S := 0$
 $\rho := 1$
 Boucle : $\sigma := \rho$
 $\rho := \|R\|^2$
 SI $\rho < \varepsilon$ alors aller à Fin
 $S := {}^T A \times R + (\rho/\sigma) \times S$
 $\sigma := -\rho / \|S\|^2$
 $X := X + \sigma \times S$
 $R := R + \sigma \times A \times S$
 Aller à Boucle
 Fin :

PROGRAMME ALGOL

PROCEDURE RESO (A) SECOND MEMBRE ET RESULTAT :
 (X)
 DIMENSION : (N) LIMITE : (EPS);
 ARRAY A, X; INTEGER N; REAL EPS;
 BEGIN ARRAY R, S [1 : N]; INTEGER I, J;
 REAL Z, RO, SIG;
 INITIALISATION : FOR I := N, I-1 WHILE I > 0 DO
 BEGIN R [I] := -X [I];
 X [I] := S [I] := 0
 END;
 RO := 1;
 BOUCLE : SIG := RO; RO := 0; FOR I := N, I-1
 WHILE I > 0 DO
 RO := RO + R [I] ↑ 2;
 IF RO > EPS THEN
 BEGIN FOR I := N, I-1 WHILE I > 0 DO
 BEGIN Z := RO/SIG × S [I];


```

FOR J := N, J-1 WHILE
  J > 0 DO
  Z := Z + A
  [J, I] × R [J];
  S [I] := Z
END;
SIG := 0; FOR I := N, I-1,
WHILE I > 0 DO
  SIG := SIG + S [I] ↑ 2;
  SIG := -RO/SIG;
FOR I := N, I-1 WHILE I > 0
DO
  BEGIN X [I] := X [I] + SIG
    × S [I]; Z := 0;
  FOR J := N, J-1
  WHILE J > 0 DO
    Z := Z + A
    [I, J] × S [J];
    R [I] := R [I]
    + SIG × Z
  END; GO TO BOUCLE
END
END;

```

COMPARAISON

Nous avons écrit un programme qui résoud un système linéaire à matrice bande, dont le second membre est calculé de telle sorte que la solution soit connue à l'avance et, successivement, par cette méthode puis par la méthode du gradient conjuguée, après multiplication par la matrice transposée.

Le programme ne tient pas compte des vides éventuels à l'intérieur de la bande, vides qui disparaissent en général lors de la multiplication matricielle. En tenant compte, il serait probablement possible de rendre la première méthode aussi rapide, dans certains cas, que la seconde.

Du point de vue rapidité, notre méthode demande pour le calcul environ 12 % de temps de plus que l'autre, mais il faut ajouter à la seconde le temps du produit matriciel.

Comme, par définition, nous connaissons la solution, nous calculons la norme euclidienne de la différence entre le vecteur itéré et le vecteur solution exacte.

Les résultats donnés ci-dessous s'entendent à nombre égal d'itérations.

n représente la dimension,

ϵ_1 la norme euclidienne de l'erreur de notre méthode,

ϵ_2 la norme euclidienne de l'erreur de l'autre méthode.

Matrices du type

$$\begin{array}{cccc}
 0 & -2 & 1 & \\
 2 & 0 & -2 & 1 \\
 -1 & 2 & 0 & -2 & 1 \\
 & -1 & 2 & 0 & -2 & 1 \\
 & & \dots\dots\dots & & & \\
 & & & -1 & 2 & 0 & -2 & 1 \\
 & & & & -1 & 2 & 0 & -2 \\
 & & & & & x & y & z
 \end{array}$$

- $x = 1, y = -2, z = 2, n = 71, \epsilon_1 = 3 \cdot 10^{-10}, \epsilon_2 = 10^{-6}$
- $x = 1, y = -2, z = 1, n = 71, \epsilon_1 = 2 \cdot 10^{-9}, \epsilon_2 = 3 \cdot 10^{-6}$
- $x = -2, y = 4, z = -1, n = 71, \epsilon_1 = 4 \cdot 10^{-7}, \epsilon_2 = 0 \cdot 1$
- $x = -4, y = 8, z = -1, n = 71, \epsilon_1 = 3 \cdot 10^{-5}, \epsilon_2 = 100$
- $x = -2, y = 4, z = -1, n = 71, \epsilon_1 = 5 \cdot 10^{-11}, \epsilon_2 = 2 \cdot 10^{-5}$
- $x = -4, y = 8, z = -1, n = 71, \epsilon_1 = 10^{-9}, \epsilon_2 = 5 \cdot 10^{-5}$

Matrices du type

$$\begin{array}{cccc}
 2 & -3 & 0 & 1 \\
 4 & -6 & 1 & 0 & 1 \\
 0 & 4 & -6 & 1 & 0 & 1 \\
 2 & 0 & 1 & -4 & -2 & 0 & 3 \\
 & & \dots\dots\dots & & & \\
 & & & 2 & 0 & 1 & -4 & -2 & 0 & 3 \\
 & & & & -3 & 0 & 2 & 6 & -7 & 0 \\
 & & & & & -3 & 0 & 2 & 6 & -7 \\
 & & & & & & 2 & 0 & -7 & 5
 \end{array}$$

$$n = 90, \quad \varepsilon_1 = 2 \cdot 10^{-6}, \quad \varepsilon_2 = 0,04$$

$$n = 115, \quad \varepsilon_1 = 8 \cdot 10^{-6}, \quad \varepsilon_2 = 500$$

Matrice telle que

$a_{i-1,i} = -3, a_{i+1,i} = 3, a_{i-3,i} = 1, a_{i+3,i} = -1$ pour tout i
 et $a_{n,n} = 1$ les autres éléments tous nuls.

$$n = 74, \quad \varepsilon_1 = 6 \cdot 10^{-4}, \quad \varepsilon_2 = 6 \cdot 10^{-2}$$

Matrice

1	1						
1	2	1					
1	0	1	2	-4	-4		
-2	0	4	2	0	-1	0	1
-2	0	4	2	0	-1	0	1
.....							
			-2	0	4	2	0 -1 0 1
							2 1 -2 1 2
							2 1 -2 1 2
							2 1 -2 1 2
							1 2

$$n = 115, \quad \varepsilon_1 = 2 \cdot 10^{-4}, \quad \varepsilon_2 = 15$$

Matrice telle que

$a_{i+k,i}$ est égal pour les différentes valeurs de k à

k	-15	-12	-11	-7	-6	-5	0	1	2	3	4
$a_{i+k,i}$	1	2	11	7	9	8	13	15	23	17	5

k	9	10	11	15	
$a_{i+k,i}$	11	19	23	19	

$$n = 95, \quad \varepsilon_1 = 2 \cdot 10^{-11}, \quad \varepsilon_2 = 2 \cdot 10^{-7}$$

$$n = 115, \quad \varepsilon_1 = 5 \cdot 10^{-11}, \quad \varepsilon_2 = 10^{-7}$$

Matrice du même type mais avec :

k	-17	-14	-13	-9	-8	-7	-2	-1	0	1	2
$a_{i+k,i}$	1	2	11	7	9	8	13	15	23	17	5

k	7	8	9	13	14	17	
$a_{i+k,i}$	11	19	23	19	47	43	

$$n = 67, \quad \varepsilon_1 = 2 \cdot 10^{-9}, \quad \varepsilon_2 = 10^{-6}$$

$$n = 115, \quad \varepsilon_1 = 4 \cdot 10^{-8}, \quad \varepsilon_2 = 1$$

BIBLIOGRAPHIE

En ce qui concerne les méthodes numériques citées et le théorème de Steifel Hestenes :

NOËL GASTINEL, *Analyse Numérique Linéaire*, Hermann, 1966.

En ce qui concerne le langage Algol :

J. W. BACKUS, F. L. BAUER, J. GREEN, C. KATZ, J. MCCARTHY, P. NAUR, A. J. PERLIS, H. RUTISHAUSER, K. SAMELSON, B. VAUQUOIS, J. H. WEGSTEIN, A. van WIJNGAARDEN et M. WOODGER, *Revised report on the algorithmic language Algol 60*, Regnecentralen, Copenhagen, 1962.

BOLLIET, GASTINEL et LAURENT, Un nouveau langage scientifique Algol 60, *Manuel Pratique*, Hermann.