Méthodes des sous-espaces de Krylov pour la résolution des systèmes linéaires cas: standard, global et par blocs

A. Messaoudi

Département de Mathématiques
Ecole Normale Supérieure,
Université Mohammed V,
Rabat, Maroc.
abderrahim.messaoudi@gmail.com

Master 2 de Mathématiques, 2017-2018

1ère PARTIE

CAS STANDARD

1 Introduction

On considère le système linéaire

$$Ax = b, \tag{1}$$

où A est une matrice carrée réelle d'ordre n inversible, $b \in \mathbb{R}^n$.

Soit x_0 un vecteur initial donné et $r_0 = b - Ax_0$ le résidu associé à x_0. La solution du système (1) peut s'écrire alors comme suit

$$x = A^{-1}b \Leftrightarrow x = x_0 + A^{-1}b - x_0 \Leftrightarrow x = x_0 + A^{-1}(b - Ax_0),$$

$$x = x_0 + A^{-1}r_0, \quad \text{avec} \quad r_0 = b - Ax_0.$$

Soit $\mathbb{P}_n$ l'ensemble des polynômes défini par

$$\mathbb{P}_n = \{p_k(t) = \sum_{i=0}^{k} \alpha_i t^i, k \leq n \text{ et } \alpha_0 = p_k(0) = 1\}.$$

Définition 1 *Polynôme minimal de A pour r_0*
$p_m \in \mathbb{P}_n$ est dit polynôme minimal de la matrice A pour le vecteur r_0 ssi $p_m(A)r_0 = 0$ avec $m = min\{k | p_k \in \mathbb{P}_n,\ p_k(A)r_0 = 0\}$.

Remarque 1

$$m \leq \mu \leq n,$$

où μ est le degré du polynôme minimal de A et n le degré du polynôme caractéristique de A.

On a alors

$$p_m(A)r_0 = \sum_{i=0}^{m} \alpha_i A^i r_0 = r_0 + \sum_{i=1}^{m} \alpha_i A^i r_0 = 0. \tag{2}$$

Si on pose

$$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_m)^T \in \mathbb{R}^m,$$

$$W_m = [Ar_0, A^2r_0, \cdots, A^m r_0] = A[r_0, Ar_0, \cdots, A^{m-1}r_0] = AK_m,$$

à partir de (2) on obtient

$$r_0 = -\sum_{i=1}^{m} \alpha_i A^i r_0 = -W_m \alpha,$$

$$x = x_0 + A^{-1}r_0 = x_0 - K_m\alpha = x_0 - \sum_{i=1}^{m} \alpha_i A^{i-1} r_0.$$

Donc

$$x - x_0 \in sev\{r_0, Ar_0, \cdots, A^{m-1}r_0\} = K_m(A, r_0),$$

où $K_m(A, r_0)$ est le sous-espace vectoriel engendré par $r_0, Ar_0, \cdots, A^{m-1}r_0$

Définition 2 *Sous-espaces de Krylov*
pour $1 \leq k \leq m$

$$K_k(A, r_0) = sev\{r_0, Ar_0, \cdots, A^{k-1}r_0\} = span\{r_0, Ar_0, \cdots, A^{k-1}r_0\},$$

est le sous-espace de Krylov d'ordre k *associé à la matrice* A *et au vecteur* r_0.

La détermination d'un élément de $K_m(A, r_0)$ nécessite le calcul des puissances de A. Il va falloir donc construire une base de $K_m(A, r_0)$ qui évite de tels calculs : le procédé d'Arnoldi.

2 Procédé d'Arnoldi : cas standard

On va d'abord rappeler la décomposition QR d'une matrice puis on donnera le procédé (ou processus) d'Arnoldi.

2.1 La décomposition QR d'une matrice

Proposition 1 *La Décomposition QR*
Soit $U = [u_1, u_2, \cdots, u_m]$ *une* $n \times m$ *matrice, avec* $m \leq n$, *de rang maximal (=m), alors il existe* $Q = [q_1, q_2, \cdots, q_m]$, *orthogonale* ($Q^TQ = I_m$) *et* $R = (r_{i,j})_{1 \leq i,j \leq m}$ *triangulaire supérieure telle que* $U = QR$.

La démonstrastion de cette proposition se fait en remarquant que

$$U = QR, \quad \Leftrightarrow \; u_i = \sum_{j=1}^{i} q_j r_{j,i}, \quad \text{pour } i = 1, \cdots, m,$$

puis on utilise le fait que la matrice Q est orthogonale ($Q^TQ = I_m$).

Pour le calcul de Q et de R on utilise le procédé de Gram-Schmidt suivant.

Algorithme 1 *Procédé de Gram-Schmidt*

$r_{1,1} = \|u_1\|_2$, $q_1 = \dfrac{u_1}{r_{1,1}}$,

pour $i = 2, \cdots, m$, *faire*

$\quad r_{j,i} = \langle u_i, q_j \rangle_2, \quad$ *pour* $j = 1, \cdots, i-1$

$\quad \tilde{u} = u_i - \displaystyle\sum_{j=1}^{i-1} r_{j,i} q_j,$

$\quad r_{i,i} = \|\tilde{u}\|_2,$

$\quad q_i = \dfrac{\tilde{u}}{r_{i,i}},$

fin du i.

Avec $\langle ., . \rangle_2$ le produit scalaire euclidien sur $\mathbb{R}^n$ et $\|.\|_2$ la norme associée.
Mais le procédé de Gram-Schmidt modifié est plus stable numériquement

Algorithme 2 *Procédé de Gram-Schmidt modifié*

$r_{1,1} = \|u_1\|_2,\ q_1 = \dfrac{u_1}{r_{1,1}},$

pour $i = 2, \cdots, m,$ *faire*

$\quad \tilde{u} = u_i,$

$\quad$ *pour* $j = 1, \cdots, i-1,$ *faire*

$\quad\quad r_{j,i} = \langle \tilde{u}, q_j \rangle_2,$

$\quad\quad \tilde{u} = \tilde{u} - r_{j,i} q_j,$

$\quad$ *fin du j,*

$\quad r_{i,i} = \|\tilde{u}\|_2,$

$\quad q_i = \dfrac{\tilde{u}}{r_{i,i}},$

fin du i.

2.2 Procédé d'Arnoldi : cas standard

Ce procédé permet de construire une base orthonormée du sous-espace de Krylov $K_k(A, v)$.

Algorithme 3 *Procédé d'Arnoldi*

$\beta = \|v\|_2,\ v_1 = \dfrac{v}{\beta},$

pour $i = 1, \cdots, k,$ *faire*

$\quad h_{j,i} = \langle Av_i, v_j \rangle_2, \quad$ *pour* $j = 1, \cdots, i$

$$\tilde{v} = Av_i - \sum_{j=1}^{i} h_{j,i} v_j,$$

$\quad h_{i+1,i} = \|\tilde{v}\|_2,$ *si* $h_{i+1,i} = 0$ *arrêter*

$\quad v_{i+1} = \dfrac{\tilde{v}}{h_{i+1,i}},$

fin du i.

Ou bien le procédé d'Arnoldi modifié

Algorithme 4 *Procédé d'Arnoldi modifié*

$\beta = \|v\|_2,\ v_1 = \dfrac{v}{\beta},$

pour $i = 1, \cdots, k,$ *faire*

$\quad \tilde{v} = Av_i,$

$\quad$ *pour* $j = 1, \cdots, i,$ *faire*

$\quad\quad h_{j,i} = \langle \tilde{v}, v_j \rangle_2,$

$\quad\quad \tilde{v} = \tilde{v} - h_{j,i} v_j,$

$\quad$ *fin du j,*

$\quad h_{i+1,i} = \|\tilde{v}\|_2,$ *si* $h_{i+1,i} = 0$ *arrêter*

$\quad v_{i+1} = \dfrac{\tilde{v}}{h_{i+1,i}},$

fin du i.

Remarque 2 *L'algorithme 4 s'arrête à l'étape* k *ssi* $h_{k+1,k} = 0$ *et* $\{v_1, v_2, \cdots, v_k\}$ *forme une base orthonormale de* $K_k(A, v)$.

Notons par $H_k = (h_{i,j})_{1\leq i,j\leq k}$ la matrice de Hessenberg supérieure (i.e. $h_{i,j} = 0$, $i-j>1$) contruite par l'Algorithme 4

$$H_k = \begin{pmatrix} h_{1,1} & h_{1,2} & h_{1,3} & \cdots & h_{1,k} \\ h_{2,1} & h_{2,2} & h_{2,3} & \ddots & \vdots \\ 0 & h_{3,2} & h_{3,3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & h_{k-1,k} \\ 0 & \cdots & 0 & h_{k,k-1} & h_{k,k} \end{pmatrix},$$

V_k est la matrice définie par

$$V_k = [v_1, v_2, \cdots, v_k],$$

et e_k le vecteur suivant

$$e_k = (0, 0, \cdots, 0, 1)^T \in \mathbb{R}^k,$$

notons aussi par $\tilde{H}_k$ la $(k+1)\times k$ matrice donnée par

$$\tilde{H}_k = \begin{pmatrix} H_k \\ h_{k+1,k}e_k^T \end{pmatrix}.$$

Les propriétés suivantes découlent immédiatement de l'Algorithme 3

Proposition 2

$$AV_k = V_{k+1}\tilde{H}_k = V_kH_k + h_{k+1,k}v_{k+1}e_k^T, \tag{3}$$

$$H_k = V_k^TAV_k, \quad \tilde{H}_k = V_{k+1}^TAV_k, \tag{4}$$

$$v_{k+1} = p_k(A)v_1. \tag{5}$$

Preuve.
A partir de l'algorithme 3 on a

$$Av_j = \sum_{i=1}^{j+1} h_{i,j}v_i, \quad \text{pour} \quad j = 1, \cdots, k,$$

soit matriciellement

$$AV_k = V_kH_k + h_{k+1,k}v_{k+1}e_k^T,$$

ou encore

$$AV_k = V_{k+1}\tilde{H}_k;$$

la relation (4) est obtenue en multipliant la relation (3) par V_k^T soit

$$V_k^TAV_k = V_k^TV_kH_k + h_{k+1,k}V_k^Tv_{k+1}e_k^T = H_k.$$

Pour démontrer la ralation (5) on procède par récurrence. ■

Proposition 3 *L'algorithme 4 s'arrête à l'étape m ($h_{m+1,m} = 0$) ssi le degré du polynôme minimal de la matrice A pour le vecteur v est égal à m ($p_m(A)v_1 = 0$).*

3 Quelques méthodes de résolution de systèmes linéaires

On va s'intéresser aux méthodes qui utilisent les sous-espaces de Krylov, donc le procédé d'Arnoldi. Parmi ces méthodes il y'a celles qu'on appelle méthodes de résidus orthogonaux (ou bien de projection oblique) et celles dites de résidu minimal (ou bien de projection orthogonale), et d'autres qui utilisent la biorthogonalité.

3.1 La méthode d'Arnoldi : Full Orthogonalization Method

Soit x_0 une approximation arbitraire de la solution $x = A^{-1}b$ et $r_0 = b - Ax_0$ le résidu associé à x_0. La méthode d'Arnoldi pour la résolution du système linéaire (1), dite FOM, définit, à l'étape k, une approximation x_k de x telle que $x_k - x_0 \in K_k(A, r_0)$ soit $x_k = x_0 + V_k y$ avec $y \in \mathbb{R}^k$ dont le résidu associé $r_k = b - Ax_k$ vérifie la condition de Galerkin

$$r_k = b - Ax_k \perp K_k(A, r_0).$$

Si l'on choisit $v_1 = r_0/\beta$ en posant $\beta = \| r_0 \|_2$ et si on utilise la procédure d'Arnoldi pour construire une base $\{v_1, \cdots, v_k\}$ du sous-espace de Krylov $K_k(A, v_1) = K_k(A, r_0)$ on aura alors $V_k A V_k = H_k$ et si H_k est inversible alors la condition

$$r_k \perp K_k(A, v_1) \quad \Leftrightarrow \quad V_k^T r_k = V_k^T r_0 - V_k^T A V_k y = 0,$$

d'où

$$y = (V_k^T A V_k)^{-1} V_k^T r_0 = \beta H_k^{-1} V_k^T v_1 = \beta H_k^{-1} e_1^{(k)},$$

avec $e_1^{(k)}$ le 1er vecteur de la base canonique de $\mathbb{R}^k$, on obtient donc

$$x_k = x_0 + V_k y = x_0 + \beta V_k H_k^{-1} e_1^{(k)}.$$

Algorithme 5 *Full Orthogonalization Method (FOM)*

1. *Choisir* x_0*, calculer* $r_0 = b - Ax_0$*,* $\beta = \| r_0 \|_2$ *et* $v_1 = r_0/\beta$
2. $H_k = (h_{i,j})$ *:* $k \times k$*,* $H_k = 0$
3. *pour* $j = 1, 2, \cdots, k$*, faire*
   - $\tilde{v} = Av_j$
   - *pour* $i = 1, \cdots, j$*, faire*
     - $h_{i,j} = (\tilde{v}, v_i)_2$
     - $\tilde{v} = \tilde{v} - h_{i,j} v_i$
   - *fin* i*,*
   - $h_{j+1,j} = \| \tilde{v} \|_2$*, si* $h_{j+1,j} = 0$ *poser* $j = k$ *et aller à 4*
   - $v_{j+1} = \tilde{v}/h_{j+1,j}$

   fin j*,*
4. *calculer* $y_k = \beta H_k^{-1} e_1^{(k)}$ *et* $x_k = x_0 + V_k y_k$.

Proposition 4 *Le vecteur résidu associé à l'approximation* x_k *donnée par l'algorithme FOM est tel que*

$$r_k = b - Ax_k = -h_{k+1,k} e_k^T y_k v_{k+1},$$

et on a

$$\| r_k \|_2 = h_{k+1,k} |e_k^T y_k|.$$

Preuve.

On a

$$\begin{aligned} r_k &= b - Ax_k = b - A(x_0 + V_k y_k) \\ &= r_0 - AV_k y_k \\ &= r_0 - (V_k H_k + h_{k+1,k} v_{k+1} e_k^T) y_k \\ &= \beta v_1 - V_k H_k y_k - h_{k+1,k} v_{k+1} e_k^T y_k \\ &= \beta v_1 - \beta V_k e_1 - h_{k+1,k} e_k^T y_k v_{k+1} \\ &= \beta v_1 - \beta v_1 - h_{k+1,k} e_k^T y_k v_{k+1}, \end{aligned}$$

et par suite

$$\| r_k \|_2 = h_{k+1,k} |e_k^T y_k|.$$

Remarque 3 *Dans l'algorithme FOM on a k multiplications matrice-vecteur et k vecteurs à stocker v_i, lorsque k devient grand les nombres de multiplications et de stokcage deviennent grands aussi. Il y'a deux façons pour eviter ce problème.*

Redémarrage : utiliser FOM pour s étapes, calculer x_s si le résultat est satisfaisant alors arrêter sinon poser $x_0 = x_s$ et redémarrer l'algorithme avec ce nouveau vecteur initial, d'où l'algorithme de FOM redémarré noté par FOM(s).

Algorithme 6 *La Méthode d'Arnoldi Redémarrée : FOM(s)*

Choisir x_0,

1. *calculer $r_0 = b - Ax_0$, $\beta = \| r_0 \|_2$ et $v_1 = r_0/\beta$*
2. *utiliser le processus d'Arnoldi modifié pour calculer H_s et (v_i), $i = 1, \cdots, s$*
3. *calculer $y_s = \beta H_s^{-1} e_1^{(s)}$ et $x_s = x_0 + V_s y_s$, si satisfait arrêter*
4. *poser $x_0 = x_s$ et aller à 1.*

Tronquage : On tronque le calcul des $\{v_j\}$: on orthogonalise v_j avec seulement les k derniers vecteurs (s un entier fixé $s \leq m$). Normalement on a

$$h_{j+1,j} v_{j+1} = Av_j - \sum_{i=1}^{j} h_{i,j} v_i,$$

et lorsqu'on tronque on aura

$$h_{j+1,j} v_{j+1} = Av_j - \sum_{i=max(1,j-s+1)}^{j} h_{i,j} v_i.$$

Attention lorsqu'on tronque la structure de H_s change.

3.2 La méthode du GMRes (Generalized Minimal Residual)

Soit x_0 une approximation initiale arbitraire de $x = A^{-1}b$, et $r_0 = b - Ax_0$ le résidu associé à x_0, soit aussi $\beta = \| r_0 \|_2$ et $v_1 = r_0/\beta$, on utilise le processus d'Arnoldi pour construire une base orthonormale $\{v_i\}$ du sous-espace de Krylov $K_k(A, v_1) = K_k(A, r_0)$ et la matrice de Hessenberg H_k. La méthode du GMRes consiste à chercher l'approximation x_k tels que $x_k - x_0 \in K_(A, r_0)$ dont la norme du résidu associé soit minimale. Soit $z-x_0 \in K_k(A,r_0)$ et $r = b-Az$ alors on aura à considérer la fonctionnelle $J(y) = \| r \|_2$. Donc $z = x_0 + V_k y$ avec $y \in \mathbb{R}^k$ et

$$J(y) = \| b - Az \|_2 = \| b - A(x_0 + V_k y) \|_2 = \| r_0 - AV_k y \|_2,$$

donc on cherche à minimiser $J(y)$ sur $x_0 + K_k(A, r_0)$ soit

$$\min_{z \in x_0 + K_k(A,r_0)} \| b - Az \|_2 = \min_{y \in \mathbb{R}^k} \| r_0 - AV_k y \|_2 .$$

Or en utilisant la relation (3) on obtient

$$\begin{aligned} J(y) &= \| b - Az \|_2 = \| r_0 - AV_k y \|_2 \\ &= \| \beta v_1 - V_{k+1} \tilde{H}_k y \|_2 \\ &= \| V_{k+1}(\beta e_1 - \tilde{H}_k y) \|_2 \\ &= \| \beta e_1 - \tilde{H}_k y \|_2, \end{aligned}$$

avec e_1 le 1er vecteur de la base canonique de $\mathbb{R}^{k+1}$, donc si

$$y_k = arg \min_{y \in \mathbb{R}^k} \| \beta e_1 - \tilde{H}_k y \|_2, \tag{6}$$

alors

$$x_k = x_0 + V_k y_k, \tag{7}$$

d'où l'algorithme du GMRes.

Algorithme 7 *La méthode du GMRes*

1. *Choisir* x_0*, calculer* $r_0 = b - Ax_0$*,* $\beta = \| r_0 \|_2$ *et* $v_1 = r_0/\beta$
2. $H_k = (h_{i,j})$ *:* $k \times k$*,* $H_k = 0$
3. *pour* $j = 1, 2, \cdots, k$*, faire*

   $\tilde{v} = Av_j$

   pour $i = 1, \cdots, j$*, faire*

   $h_{i,j} = (\tilde{v}, v_i)_2$

   $\tilde{v} = \tilde{v} - h_{i,j} v_i$

   fin i*,*

   $h_{j+1,j} = \| \tilde{v} \|_2$*, si* $h_{j+1,j} = 0$ *poser* $j = k$ *et aller à 4*

   $v_{j+1} = \tilde{v}/h_{j+1,j}$

   fin j*,*
4. *calculer* y_k *solution de (6) et calculer* x_k *par la relation (7).*

On va montrer que l'itéré x_k calculé par la relation (7) avec y_k solution du problème (6) est équivalent à trouver $x_k \in x_0 + K_k(A, r_0)$ tel que $r_k = b - Ax_k \perp AK_k(A, r_0)$. La solution du problème (6) est donnée par ($\tilde{H}_k$ est rectangulaire)

$$y_k = \beta \tilde{H}_k^+ e_1, \text{ avec } \tilde{H}_k^+ = \left(\tilde{H}_k^T \tilde{H}_k \right)^{-1} \tilde{H}_k^T, \text{ est le pseudo-inverse de } \tilde{H}_k.$$

Soit en utilisant la relation (3) ($AV_k = V_{k+1}\tilde{H}_k$), le fait que V_k est orthonormale et que $V_{k+1}^T(r_0) = \beta e_1$, on obtient

$$\begin{aligned}
y_k &= \beta \tilde{H}_k^+ e_1 \\
&= \left(\tilde{H}_k^T \tilde{H}_k\right)^{-1} \tilde{H}_k^T(\beta e_1) \\
&= \left(\tilde{H}_k^T V_{k+1}^T V_{k+1} \tilde{H}_k\right)^{-1} \tilde{H}_k^T V_{k+1}^T r_0 \\
&= \left[(V_{k+1}\tilde{H}_k)^T V_{k+1}\tilde{H}_k\right]^{-1} (V_{k+1}\tilde{H}_k)^T r_0 \\
&= \left[(AV_k)^T AV_k\right]^{-1} (AV_k)^T r_0.
\end{aligned}$$

Maintenant soit $x_k = x_0 + V_k z_k$ tel que le résidu associé à x_k

$$r_k = b - Ax_x = r_0 - AV_k z_k \perp AK_k(A, r_0)$$

i.e. $(Av_i)^T r_k = 0$, pour $i = 1, \cdots, k$, soit matriciellement $(AV_k)^T r_k = 0$, ainsi on obtient lorsque la matrice $(AV_k)^T AV_k$ est inversible

$$(AV_k)^T r_k = (AV_k)^T (r_0 - AV_k z_k) = 0, \ \Leftrightarrow \ z_k = \left[(AV_k)^T AV_k\right]^{-1} (AV_k)^T r_0 = y_k.$$

Donc pour la méthode du GMRes on a

$$\begin{aligned}
x_k &= x_0 + V_k y_k \\
&= x_0 + \beta V_k \tilde{H}_k^+ e_1 \\
&= x_0 + V_k \left[(AV_k)^T AV_k\right]^{-1} (AV_k)^T r_0.
\end{aligned}$$

Ainsi on a la proposition suivante.

Proposition 5 *La méthode du GMRes peut être définie, à l'étape k, par*

$$\begin{cases}
x_k = x_0 + V_k y_k \\
\\
y_k = arg \min\limits_{y \in \mathbb{R}^k} \parallel \beta e_1 - \tilde{H}_k y \parallel_2,
\end{cases}$$

ou bien par

$$\begin{cases}
x_k = x_0 + V_k z_k \\
\\
r_k = b - Ax_k \perp AK_k(A, r_0)
\end{cases}$$

Proposition 6 *La matrice $(AV_k)^T AV_k$ est inversible et la méthode du GMRes ne présente pas de break-down (division par zéro).*

Preuve.

V_k est une matrice de rang maximale (=k) car $\{v_i\}_{i=1,\ldots,k}$ est une base de $K_k(A, r_0)$, la matrice A est supposée inversible donc AV_k est une matrice de rang maximal et par conséquent $(AV_k)^T AV_k$ est inversible car elle est carrée et de rang maximal. ■

Pour l'implémentation de l'algorithme 7 on va utiliser la décomposition QR pour résoudre le problème (6), pour ceci on va utiliser les rotations de Givens pour décomposer $\tilde{H}_k$. Soit $\Omega_i = (\omega_{i,j})_{1\le i,j\le k+1}$ la ième rotation de Givens seuls les éléments

$$\omega_{(i,i)} = c_i,\ \omega_{(i,i+1)} = s_i,\ \omega_{(i+1,i)} = -s_i,\ \omega_{(i+1,i+1)} = c_i$$

peuvent être non nuls

$$\Omega_i = \begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & 0 & & & \\ & & 1 & & & & & \\ & & & c_i & s_i & & & \\ & & & -s_i & c_i & & & \\ & & & & & 1 & & \\ & & & 0 & & & \ddots & \\ & & & & & & & 1 \end{pmatrix}, \quad \text{avec} \quad s_i^2 + c_i^2 = 1.$$

Remarquons que $\Omega_i^T\Omega_i = \Omega_i\Omega_i^T = I_{k+1}$. Soit $\tilde{H}_k = (h_{i,j})_{1\le i\le k+1,\ 1\le j\le k}$

$$\tilde{H}_k = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & \cdots & h_{1,k} \\ h_{2,1} & h_{2,2} & \cdots & \cdots & h_{2,k} \\ 0 & h_{3,2} & h_{3,3} & \cdots & h_{3,k} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & h_{k,k-1} & h_{k,k} \\ 0 & \cdots & \cdots & 0 & h_{k+1,k} \end{pmatrix}.$$

On va définir $\Omega_1, \cdots, \Omega_k$ tel que

$$\Omega_k \times \Omega_{k-1} \times \cdots \times \Omega_1 \times \tilde{H}_k = \begin{pmatrix} \ddots & \cdots & \cdots & \cdots \\ & \ddots & \cdots & \cdots \\ & & \ddots & \cdots \\ & & & \ddots \\ \hline 0 & \cdots & \cdots & 0 \end{pmatrix} = \bar{R}_k.$$

Posons $\tilde{H}_k^{(0)} = \tilde{H}_k$ et pour $m = 1, \cdots, k$, $\tilde{H}_k^{(m)} = \Omega_m \tilde{H}_k^{(m-1)} = (h_{i,j}^{(m)})_{1\le i\le k+1,\ 1\le j\le k}$.
On aura pour les premières étapes :

$$\Omega_1 = \begin{pmatrix} c_1 & s_1 & & & \\ -s_1 & c_1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}, \quad \text{avec} \quad c_1 = \frac{h_{1,1}}{\sqrt{h_{1,1}^2 + h_{2,1}^2}}, \quad s_1 = \frac{h_{2,1}}{\sqrt{h_{1,1}^2 + h_{2,1}^2}}.$$

D'où

$$\Omega_1 \times \tilde{H}_k = \tilde{H}_k^{(1)} = \begin{pmatrix} h_{1,1}^{(1)} & h_{1,2}^{(1)} & \cdots & \cdots & h_{1,k}^{(1)} \\ 0 & h_{2,2}^{(1)} & \cdots & \cdots & h_{2,k}^{(1)} \\ 0 & h_{3,2} & h_{3,3} & \cdots & h_{3,k} \\ \vdots & \ddots & \ddots & \cdots & \vdots \\ \vdots & & \ddots & h_{k,k-1} & h_{k,k} \\ 0 & \cdots & \cdots & 0 & h_{k+1,k} \end{pmatrix}.$$

A la deuxième étape on aura

$$\Omega_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & c_2 & s_2 & 0 & \cdots & 0 \\ 0 & -s_2 & c_2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \quad c_2 = \frac{h_{2,2}^{(1)}}{\sqrt{(h_{2,2}^{(1)})^2 + h_{3,2}^2}}, \quad s_2 = \frac{h_{3,2}}{\sqrt{(h_{2,2}^{(1)})^2 + h_{3,2}^2}}.$$

On obtient

$$\Omega_2 \times \Omega_1 \times \tilde{H}_k = \tilde{H}_k^{(2)} = \begin{pmatrix} h_{1,1}^{(1)} & h_{1,2}^{(1)} & \cdots & \cdots & h_{1,k}^{(1)} \\ 0 & h_{2,2}^{(2)} & \cdots & \cdots & h_{2,k}^{(2)} \\ 0 & 0 & h_{3,3}^{(2)} & \cdots & h_{3,k}^{(2)} \\ \vdots & \vdots & h_{4,3} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & h_{k,k} \\ 0 & \cdots & \cdots & \cdots & h_{k+1,k} \end{pmatrix}.$$

Ainsi de suite à l'étape k on obtient

$$\Omega_k \times \cdots \times \Omega_1 \tilde{H}_k = \tilde{H}_k^{(k)} = \begin{pmatrix} h_{1,1}^{(k)} & h_{1,2}^{(k)} & \cdots & \cdots & h_{1,k}^{(k)} \\ 0 & h_{2,2}^{(k)} & \cdots & \cdots & h_{2,k}^{(k)} \\ 0 & 0 & \ddots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ \vdots & \vdots & \cdots & \cdots & h_{k,k}(k) \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}.$$

On pose ensuite

$$Q_k = \Omega_k \times \cdots \times \Omega_1, \quad \text{et} \quad \bar{R}_k = Q_k \tilde{H}_k = \tilde{H}_k^{(k)}. \tag{8}$$

Posons aussi

$$\bar{g}_k = Q_k(\beta e_1) = (\gamma_1,\ \gamma_2,\ \cdots,\ \gamma_{k+1})^T. \tag{9}$$

Comme Q_k est unitaire ($Q_k^T Q_k = Q_k Q_k^T = I_{k+1}$) alors on obtient

$$\begin{aligned} \min_{y \in \mathbb{R}^k} \| \beta e_1 - \tilde{H}_k y \|_2 &= \min_{y \in \mathbb{R}^k} \| Q_k(\beta e_1 - \tilde{H}_k y) \|_2 \\ &= \min_{y \in \mathbb{R}^k} \| Q_k(\beta e_1) - Q_k \tilde{H}_k y \|_2 \\ &= \min_{y \in \mathbb{R}^k} \| \bar{g}_k - \bar{R}_k y \|_2 . \end{aligned}$$

Proposition 7 *Soient $\bar{R}_k$ la matrice triangulaire supérieure définie par* (8) *et $\bar{g}_k$ le vecteur défini par* (9)*. Alors on a*

1. *Le rang de AV_k est égal au rang de R_k (la matrice obtenue à partir de $\bar{R}_k$ en supprimant la dérnière ligne). En particulier si $r_{k,k} = 0$ alors A est singulière.*
2. *Le vecteur y_k qui minimise $\| \beta e_1 - \tilde{H}_k y \|_2$ est donné par*
$$y_k = R_k^{-1} g_k, \quad \text{où} \quad g_k = (\gamma,\ \gamma_2,\ \cdots,\ \gamma_k)^T.$$
3. *Le résidu à l'étape k satisfait*
$$r_k = b - Ax_k = V_{k+1}(\beta e_1 - \tilde{H}_k y_k) = V_{k+1} Q_k^T \gamma_{k+1} e_{k+1}. \tag{10}$$
et on a
$$\| r_k \|_2 = \| b - Ax_k \|_2 = |\gamma_{k+1}|. \tag{11}$$

Preuve.

1. On utilise la relation (3) on a
$$\begin{aligned} AV_k &= V_{k+1}\tilde{H}_k \\ &= V_{k+1} Q_k^T Q_k \tilde{H}_k \\ &= V_{k+1} Q_k^T \bar{R}_k, \end{aligned}$$
comme $V_{k+1}Q_k^T$ est une matrice orthogonale donc le rang de AV_k est celui de $\bar{R}_k$. Si $r_{k,k} = 0$ et comme R_k est tyriangulaire supérieure donc le rang de R_k est inférieur ou égal à $k-1$ donc ainsi que celui de AV_k et comme V_k est de rang maximal donc A doit être singulière.
2. On a
$$\begin{aligned} \| \beta e_1 - \tilde{H}_k y \|_2^2 &= \| Q_k(\beta e_1 - \tilde{H}_k y) \|_2^2 \\ &= \| Q_k(\beta e_1) - Q_k \tilde{H}_k y \|_2^2 \\ &= \| \bar{g}_k - \bar{R}_k y \|_2^2 \\ &= |\gamma_{k+1}| + \| g_k - R_k y \|_2^2; \end{aligned}$$
et en passant au minimum on obtient
$$\min_{y \in \mathbb{R}^k} \| g_k - R_k y \|_2 = 0, \text{ car } R_k \text{ est inversible donc } y_k = R_k^{-1} g_k.$$
3. On a $x = x_0 + V_k y$ et
$$\begin{aligned} b - Ax &= V_{k+1}(\beta e_1 - \tilde{H}_k y) \\ &= V_{k+1} Q_k^T Q_k(\beta e_1 - \tilde{H}_k y) \\ &= V_{k+1} Q_k^T(\bar{g}_k - \bar{R}_k y); \end{aligned}$$
$x_k = x_0 + V_k y_k$ est le vecteur qui minimise
$$\| r \|_2 = \| b - Ax \|_2 = \| V_{k+1} Q_k^T (\bar{g}_k - \bar{R}_k y) \|_2 = \| \bar{g}_k - \bar{R}_k y \|_2,$$

or $y_k = R_k^{-1} g_k$ donc

$$\begin{aligned}
\bar{g}_k - \bar{R}_k y_k &= \begin{pmatrix} g_k \\ \gamma_{k+1} \end{pmatrix} - \begin{pmatrix} R_k \\ 0^T \end{pmatrix} y_k \\
&= \begin{pmatrix} g_k \\ \gamma_{k+1} \end{pmatrix} - \begin{pmatrix} g_k \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 0^T \\ \gamma_{k+1} \end{pmatrix} \\
&= \gamma_{k+1} e_{k+1};
\end{aligned}$$

d'où

$$r_k = b - Ax_k = V_{k+1} Q_k^T (\bar{g}_k - \bar{R}_k y_k = V_{k+1} Q_k^T (\gamma_{k+1} e_{k+1}),$$

et

$$\| r_k \|_2 = \| b - Ax_k \|_2 = \| V_{k+1} Q_k^T (\gamma_{k+1} e_{k+1}) \|_2 = |\gamma_{k+1}|. \quad \blacksquare$$

Remarque 4 *Si on pose* $\bar{g}_0 = \beta e_1 \in \mathbb{R}^{k+1}$ *et pour* $m = 1, \cdots, k$

$$\bar{g}_m = \Omega_m \bar{g}_{m-1} = \left(\gamma_1^{(m)}, \cdots, \gamma_{k+1}^{(m)} \right)^T \in \mathbb{R}^{k+1},$$

alors l'équation $\bar{g}_m = \Omega_m \bar{g}_{m-1}$ *donne*

$$\gamma_{m+1}^{(m)} = -s_m \gamma_m^{(m-1)}. \tag{12}$$

Proposition 8 *Supposons que la matrice* A *est inversible, alors l'algorithme du GMRes s'arrête à l'étape* m *i.e.* $h_{m+1,m} = 0$ *si et seulement si l'approximation* x_m *est solution du système* (1).

Preuve.

La condition est nécessaire : en effet si $h_{m+1,m} = 0$ ceci implique que $s_m = 0$ et par conséquent $\gamma_{m+1} = 0$, d'après (12), donc $\| r_m \|_2 = 0$ en utilisant (11) et $r_m = b - Ax_m = 0$ et comme A est inversible alors $x_m = A^{-1} b$. La condition est suffisante : si $x_m = A^{-1} b$ ceci implique que $r_m = b - Ax_m = 0$ donc $\| r_m \|_2 = |\gamma_{m+1}| = 0$ donc $s_m = 0$ et par conséquent $h_{m+1,m} = 0$. $\blacksquare$

Travail personnel : donner l'algorithme du GMRes en utilisant les rotations de Givens pour résoudre le problème (6) et calculant x_k en fonction de x_{k-1}.