GMRES complete and partial stagnation

Gérard MEURANT

May 2011

1. GMRES
2. APS parametrization
3. Complete stagnation
4. Partial stagnation
5. A sufficient condition for non-stagnation
6. Stagnation for $n \leq 4$

GMRES

We solve

$$Ax = b$$

with GMRES (Saad and Schultz, 1986)

$$x^k = x^0 + V_k z^k$$

where V_k is the matrix whose columns are the orthogonal Arnoldi vectors and z^k is computed to minimize the norm of $r^k = b - Ax^k$
The Krylov space of order k is

$$\mathcal{K}_k(r^0, A) \equiv span\{r^0, Ar^0, \ldots, A^{k-1}r^0\}$$

Here we take $x^0 = 0, \|b\| = 1$

We assume A of order n is real and non-derogatory, b is real and GMRES terminates at iteration n

GMRES stagnation

Partial stagnation is defined as having

$$\|r^k\| < \|r^{k-1}\|,\ k = 1, \ldots, m, \quad \|r^k\| = \|r^{k-1}\|, k = m{+}1, \ldots, m{+}p{-}1,$$

$$\|r^k\| < \|r^{k-1}\|,\ k = m + p, \ldots, n$$

The norms of the residual stay the same for p iterations starting from $k = m$

Complete stagnation corresponds to $m = 0$ and $p = n$

$$\|r^k\| = \|r^0\|,\ k = 1, \ldots, n - 1, \quad \|r^n\| = 0$$

Complete stagnation

Complete stagnation of GMRES has been studied by Zavorin, O'Leary and Elman (2003) and in Zavorin's thesis (2001)

Simoncini and Szyld (2008) and Simoncini (2010) studied conditions for non–stagnation

Liesen and Tichý (2004) considered the study of worst-case GMRES for normal matrices

See also Arioli (2010)

We will use the framework of Strakoš and his co-authors who characterized the class of matrices and rhs giving the same residual norm convergence curve

The APS parametrization

Greenbaum and Strakoš (1994) proved that any convergence curve for the residual norm that can be generated with GMRES can be obtained with a matrix having prescribed eigenvalues

Greenbaum, Pták and Strakoš (1996) showed that any nonincreasing sequence of residual norms can be given by GMRES

Arioli, Pták and Strakoš (1998) gave a complete parametrization of all pairs $\{A, b\}$ generating a prescribed residual norm convergence curve

Theorem (APS)

Assume we are given n positive numbers

$$f(0) \geq f(1) \geq \cdots \geq f(n-1) > 0$$

and n complex numbers $\lambda_1, \ldots, \lambda_n$ all different from 0. Let A be a matrix of order n and b an n-dimensional vector. The following assertions are equivalent :

1- The spectrum of A is $\{\lambda_1, \ldots, \lambda_n\}$ and GMRES applied to A and b yields residuals $r^j, \; j = 0, \ldots, n-1$ such that

$$\|r^j\| = f(j), \quad j = 0, \ldots, n-1$$

2- The matrix A is of the form $A = WYCY^{-1}W^*$ and $b = Wh$, where W is a unitary matrix, Y is given by

$$Y = \begin{pmatrix} h & \begin{matrix} R \\ 0 \end{matrix} \end{pmatrix}$$

R being any nonsingular upper triangular matrix of order $n-1$, h a vector such that

$$h = (\eta_1, \ldots, \eta_n)^T, \quad \eta_j = (f(j-1)^2 - f(j)^2)^{1/2}$$

and C is the companion matrix corresponding to the polynomial q,

$$q(z) = (z - \lambda_1) \cdots (z - \lambda_n) = z^n + \sum_{j=0}^{n-1} \alpha_j z^j$$

For partial stagnation we have $\eta_j = 0, j = m+1, \ldots, m+p-1$ and for complete stagnation $\eta_j = 0$, $j = 1, \ldots, n-1$ and $h = e^n$

Some properties of GMRES

$$K = \begin{pmatrix} b & Ab & A^2 b & \cdots & A^{n-1} b \end{pmatrix} = VU = WY$$

where $V = V_n$ and the matrix U is upper triangular with positive diagonal entries

At the end we have $AV = VH$ where H is upper Hessenberg

Theorem

The matrix U^T is the Cholesky factor of the matrix $Y^T Y$. The Hessenberg matrix H is given by

$$H = UCU^{-1}$$

A QR factorization of H is $H = Q\mathcal{R}$ where $Q = V^T W$ is upper Hessenberg orthogonal and $\mathcal{R}$ is upper triangular. Moreover, we have

$$Q = UY^{-1} = U^{-T} Y^T, \quad \mathcal{R} = YCU^{-1}$$

The matrices Q and $\mathcal{R}$ are also related to the APS parametrization by $\mathcal{H} = \mathcal{R}Q = YCY^{-1}$

What is Q?

Lemma

Let $\hat{h}$ be the vector of the first $n-1$ components of h. For the inverse of the matrix Y, we have

$$Y^{-1} = \begin{pmatrix} 0 & \cdots & 0 & 1/\eta_n \\ & R^{-1} & & -R^{-1}\hat{h}/\eta_n \end{pmatrix}$$

Let $\hat{L}$ be the Cholesky factor of $R^T R - R^T \hat{h}\hat{h}^T R$. Then

$$U = \begin{pmatrix} 1 & \hat{h}^T R \\ 0 & \\ \vdots & \hat{L}^T \\ 0 & \end{pmatrix}$$

Then

$$Q = \begin{pmatrix} \hat{h}^T & \frac{1}{\eta_n} - \frac{\|\hat{h}\|^2}{\eta_n} \\ \hat{L}^T R^{-1} & -\frac{\hat{L}^T R^{-1} \hat{h}}{\eta_n} \end{pmatrix}$$

But

$$\frac{1}{\eta_n} - \frac{\|\hat{h}\|^2}{\eta_n} = \eta_n$$

Moreover $\hat{L}\hat{L}^T = R^T R - R^T \hat{h}\hat{h}^T R$. This implies

$$(R^{-T}\hat{L})(\hat{L}^T R^{-1}) = I - \hat{h}\hat{h}^T = \check{L}\check{L}^T$$

However, $R^{-T}\hat{L}$ is not necessarily the Cholesky factor because of the signs of the diagonal entries

Theorem

Let $\check{L}$ be the lower triangular Cholesky factor of the matrix $I - \hat{h}\hat{h}^T$ and $\mathcal{S}$ be a diagonal matrix whose diagonal entries are ± 1 according to the signs of the diagonal entries of R. Then

$$Q = \begin{pmatrix} \hat{h}^T & \eta_n \\ \mathcal{S}\check{L}^T & -\frac{\mathcal{S}\check{L}^T\hat{h}}{\eta_n} \end{pmatrix}$$

Moreover, the entries of $\check{L}^T$ for $j \geq i$ are

$$(\check{L}^T)_{i,j} = -\frac{\eta_i \eta_j}{\sqrt{\eta_{i+1}^2 + \cdots + \eta_n^2}\sqrt{\eta_i^2 + \cdots + \eta_n^2}},$$

$$(\check{L}^T)_{i,i} = \frac{\sqrt{\eta_{i+1}^2 + \cdots + \eta_n^2}}{\sqrt{\eta_i^2 + \cdots + \eta_n^2}}$$

Complete stagnation

We have $h = e^n$. Therefore $\hat{h} = 0$ and $\check{L} = I$

Theorem

In case of complete stagnation, the orthogonal matrix Q in $H = Q\mathcal{R}$ is

$$Q = \begin{pmatrix} 0 & 1 \\ \mathcal{S} & 0 \end{pmatrix}$$

where $\mathcal{S}$ is a diagonal matrix with ± 1 as diagonal entries according to the signs of the diagonal entries of R

Note also that since $h = e^n$, we have

$$b^T K = b^T W Y = (e^n)^T Y = (e^1)^T$$

Therefore, we have $b^T A^j b = 0,\ j = 2, \ldots, n-1$, see Zavorin, O'Leary and Elman

Theorem

We have complete stagnation in GMRES if and only if the non-derogatory real matrix A can be written as $A = W\mathcal{U}PW^T$, where W is orthogonal, $\mathcal{U}$ is upper triangular and nonsingular and

$$P = \begin{pmatrix} 0 & 1 \\ I & 0 \end{pmatrix}$$

is a permutation matrix. The right-hand side giving stagnation is $b = We^n$

The matrix Q can be written as $Q = \check{\mathcal{S}}P$. We have $A = W\mathcal{R}QW^T$ and the sign matrix $\check{\mathcal{S}}$ can be absorbed in the upper triangular matrix by defining $\mathcal{U} = \mathcal{R}\check{\mathcal{S}}$. Moreover $b = Wh = We^n$ in the APS parametrization

If we have complete stagnation for A and b, we have also stagnation for A^T and b. Moreover we also have the factorization $A = VQ\mathcal{R}V^T$

Theorem

We have complete stagnation in GMRES if and only the non-derogatory real matrix A and the real right-hand side b can be written as in one of the four cases :

$$
\begin{aligned}
A &= W\mathcal{U}PW^T, \quad b = We^n \\
A &= VP\mathcal{U}V^T, \quad b = Ve^1 \\
A &= WP^T\mathcal{L}W^T, \quad b = We^n \\
A &= V\mathcal{L}P^TV^T, \quad b = Ve^1
\end{aligned}
$$

where W and V are orthogonal matrices, $\mathcal{U}$ is an upper triangular matrix, $\mathcal{L}$ is a lower triangular matrix and P was defined before

A is orthogonally similar to one of the 4 cases :

$$\begin{pmatrix} 0 & 0 & 0 & x \\ x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \end{pmatrix}, \quad \begin{pmatrix} x & x & x & x \\ x & x & x & 0 \\ 0 & x & x & 0 \\ 0 & 0 & x & 0 \end{pmatrix}$$

$$\begin{pmatrix} x & x & 0 & 0 \\ x & x & x & 0 \\ x & x & x & x \\ x & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & x & 0 & 0 \\ 0 & x & x & 0 \\ 0 & x & x & x \\ x & x & x & x \end{pmatrix}$$

Partial stagnation

Lemma

In case of partial stagnation of GMRES the columns $j = m+1, \dots, m+p-1$, of Q are zero except for the subdiagonal entries $(j+1, j)$ which are ± 1. The rows $i = m+2, \dots, m+p$, are zero for columns i to n

Theorem

We have partial stagnation in GMRES for iterations $m \geq 0$ to $m+p-1$ if and only the non-derogatory real matrix A and the real right-hand side b can be written as

$$A = W\mathcal{R}QW^T, \quad b = WQ^Te^1$$

where W is orthogonal, $\mathcal{R}$ is upper triangular, Q is orthogonal with the sparsity structure defined in the lemma. We have also the same partial stagnation if and only if $A = VQ\mathcal{R}V^T$ and $b = Ve^1$ with V orthogonal and Q and $\mathcal{R}$ as before

A sufficient condition for non-stagnation

We have complete stagnation if and only if

$$b^T(A^j + (A^j)^T)b = 0,\ j = 1, \ldots, n-1, \quad b^T b = 1,$$

Given A, this is a polynomial system for the components of b

Theorem

Let A be a real matrix of order n and $A_j = A^j + (A^j)^T$, $j = 1, \ldots, n-1$. A sufficient condition for the non-existence of real stagnation vectors b is that there exist real scalars μ_j, $j = 1, \ldots, n-1$ such that the matrix

$$A_\mu = \sum_{j=1}^{n-1} \mu_j A_j$$

is (positive or negative) definite

The converse of this theorem is true for $n = 3, 4$

The case $n = 3$

We have the polynomial system

$$b^T(A + A^T)b = 0, \; b^T(A^2 + (A^2)^T)b = 0, \; b^T b = 1$$

This is the simultaneous annealing of 2 quadratic forms

This problem has been studied in the literature at least since Finsler in 1937
See also Bliss (1938), Reid (1938), Albert (1938), Hestenes and McShane (1940), Dines (1941,1942,1943), Donoghue (1957), Calabi (1964), Hestenes (1968), Uhlig (1973), Marcus (1978), Tsing and Uhlig (1991), Polyak (1998)

This problem is linked to the joint field of values (joint numerical range)

$$\mathcal{F}_K(A_1, A_2) = \{(x, A_1 x), (x, A_2 x), \; x \in K^n, \; \|x\| = 1\}$$

A_i symmetric matrix

The main result is

Theorem (Uhlig)

Let A_1 and A_2 be real symmetric matrices of order $n \geq 3$ and $\mathcal{Q}_i,\ i = 1,2$ be the set $\{x \in \mathbb{R}^n \,|\, x^T A_i x = 0\}$. Then the following statements are equivalent :

(i) $\exists$ *real* λ *and* μ *such that* $\lambda A_1 + \mu A_2$ *is definite*

(ii) $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \{0\}$

(iii) *trace*(YA_1) = *trace*(YA_2) *for* Y *symmetric positive semi-definite implies that* $Y = 0$

Theorem

Let A be a real matrix of order $n = 3$. There is no real stagnation vector b if and only if $\exists$ real numbers λ and μ such that $\lambda(A + A^T) + \mu(A^2 + (A^2)^T)$ is definite

The boundary of such a region (if it exists) has been characterized by Uhlig (1973)

There exist algorithms to compute a pair (λ, μ), see Crawford and Moon (1983), Higham, Tisseur and Van Dooren (2002)

A graphical interpretation

$$b^T A_1 b = 0, \quad b^T A_2 b = 0, \quad b^T b = 1$$

Assume that A_1 has one negative eigenvalue. If we diagonalize A_1 we have

$$\lambda_1 x^2 + \lambda_2 y^2 - \lambda_3 z^2 = 0$$

in the eigenvector frame with $\lambda_i > 0,\ i = 1, 2, 3$

This is the equation of a cone with an elliptical section whose axis is the z-axis. We are interested in the intersection of this cone with the unit sphere

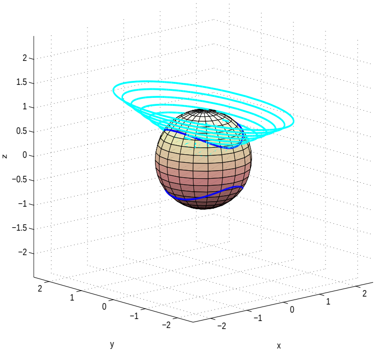
$$(\lambda_1 + \lambda_3)x^2 + (\lambda_2 + \lambda_3)y^2 = \lambda_3, \quad z^2 = 1 - x^2 - y^2$$

The intersection is the union of two smooth closed curves on the surface of the sphere

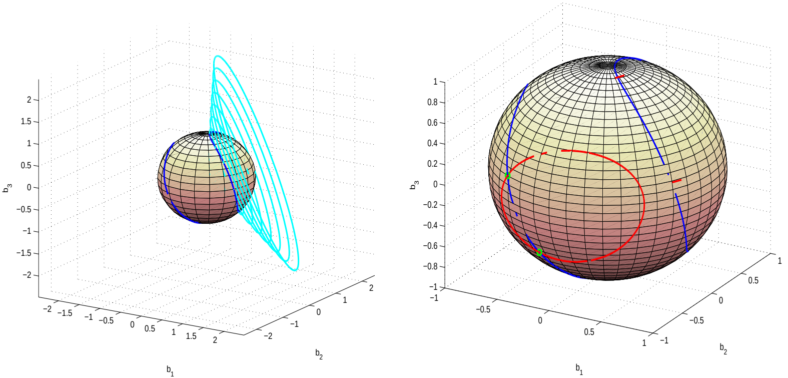
Then we do the same for A_2 and look for the intersections of the curves on the unit sphere

Examples

$$A = \begin{pmatrix} 0.614463 & 0.591283 & -1.00912 \\ 0.507741 & -0.643595 & -0.0195107 \\ 1.69243 & 0.380337 & -0.0482208 \end{pmatrix}$$



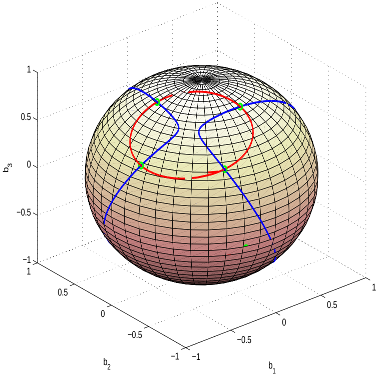
Quad. form in x, y, z frame, matrix A_1



Quad. form in b_1, b_2, b_3 frame, matrix A_1 and solutions

Here we have 4 real solutions (green circles)
The number of solutions depends on the eigenvalues of A_1 and A_2

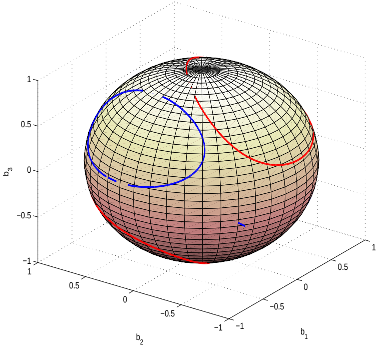
$$A = \begin{pmatrix} -0.0786619 & -1.23435 & 0.0558012 \\ -0.681657 & 0.288807 & -0.367874 \\ -1.02455 & -0.429303 & -0.464973 \end{pmatrix}$$



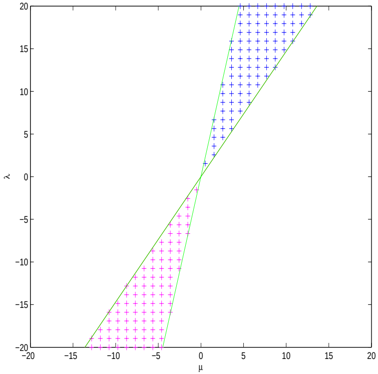
Solutions (green circle)

We have 8 real solutions

$$A = \begin{pmatrix} -0.265607 & 0.986337 & 0.234057 \\ -1.18778 & -0.518635 & 0.0214661 \\ -2.20232 & 0.327368 & -1.00394 \end{pmatrix}$$



There are no solutions



Definiteness of $\lambda A_1 + \mu A_2$, positive definite (magenta), negative definite (blue)

The case $n = 4$

Theorem (Theorem 2.1 of Polyak (1998))

Let A_1, A_2 and A_3 be real symmetric matrices of order $n \geq 3$ and $\mathcal{Q}_i,\ i = 1, 2, 3$ be the set $\{x \in \mathbb{R}^n \,|\, x^T A_i x = 0\}$, then the following statements are equivalent :

(i) *there exist μ_1, μ_2, μ_3 such that $\mu_1 A_1 + \mu_2 A_2 + \mu_3 A_3$ is positive definite*

(ii) $\mathcal{Q}_1 \cap \mathcal{Q}_2 \cap \mathcal{Q}_3 = \{0\}$ *and the set* $\{(x^T A_1 x,\ x^T A_2 x,\ x^T A_3 x), x \in \mathbb{R}^n\}$ *is an acute closed convex cone in* $\mathbb{R}^3$

For $n = 4$ we have the polynomial system

$$b^T(A + A^T)b = 0, \; b^T(A^2 + (A^2)^T)b = 0, \; b^T(A^3 + (A^3)^T)b = 0,$$

$$b^T b = 1$$

Theorem

Let A be a real matrix of order $n = 4$. Then there is no real stagnation vector if and only if there exist real μ_i, $i = 1, 2, 3$ such that $\mu_1(A + A^T) + \mu_2(A^2 + (A^2)^T) + \mu_3(A^3 + (A^3)^T)$ is definite

Let us end with an open problem about the theorem giving a sufficient condition for the non-existence of real stagnation vectors :

Under which conditions is the converse of the theorem true for $n > 4$?

Happy birthday, Paul