FM 5012
dan

The Generalized Minimal Residual (GMRES) algorithm for solving non-symmetric linear systems. *

This method was introduced by Saad and Schultz in 1986 to solve iteratively matrix equations

$$Ax = b,$$

where the matrix A is not symmetric.

(1) Motivation: the method of conjugate gradients as a Galerkin method.

We have seen that if the k-th iterate of the CG method is x^k, the error, $e^k = x - x^k$ satisfies the equality

$$w^T A e^k = 0 \qquad \forall \, w \in K_k,$$

where

$$K_k = span \{A^i r_0\}_{i=0}^{k-1}, \; r_0 = Ax_0 - b.$$

* From: Iterative methods for sparse linear systems, Y. Saad.

We immediately see that the above equations states that the k-th residual,

$$r_k = Ax_k - b,$$

is orthogonal to K_k since

$$\begin{aligned} w^T r_k &= w^T(Ax_k - b) \\ &= w^T(Ax_k - Ax) \\ &= -w^T A e_k \\ &= 0 \qquad\qquad \forall\, w \in K_k. \end{aligned}$$

Now, let $\{v_i\}_{i=1}^{k}$ be an l_2-orthonormal basis of K_k and set

$$V_k = (v_1, v_2 \dots \ v_k).$$

V_k is an $N \times k$ matrix, where $N =$ order of A. With this notation, we can write that

$$x_k = x_0 + V_k y_k,$$

where $y_k \in \mathbb{R}^k$ is the solution of

$$(V_k^T A V_k)\ y_k = -V_k^T r_0$$

Note that the above equations defines a galerkin approximation, $V_k y_k$, in K_k to the solution of

$$Az = -r_0.$$

Of course, we have that $x = x_0 + z$!
In other words, the iterates of the CG method are of the form $x_k = x_0 + z_k$, where z_k is the A-projection of z into the space K_k.

Note that if we set

$$v_1 = r_0 / \|r_0\|,$$

we have

$$\begin{aligned} x_k &= x_0 + V_k y_k, \\ y_k &= -\|r_0\| H_k^{-1} e_1, \\ H_k &= V_k^T A V_k . \end{aligned}$$

this is a rewriting of the CG method that is possible to use for matrices that are not symmetric and positive definite. Note that if k is small, the matrix H_k is easy to invert!

<u>Arnoldi's orthogonalization</u>. (of K_k).

We can obtain the basis $\{v_i\}_{i=1}^{k}$ of the Krylov subspace K_k as follows:

```
(1)  $\hat{v}_1 = r_0, \quad v_1 = \hat{v}_1 / \|\hat{v}_1\|.$

(2)  For $j = 1, \dots, k-1$ :

         $h_{ij} = v_i^T A v_j \qquad i = 1, \dots, j$
(3)      $\hat{v}_{j+1} = A v_j - \sum_{i=1}^{j} h_{ij} v_i$

(4)      $h_{j+1,j} = \|\hat{v}_{j+1}\|, \quad v_{j+1} = \hat{v}_{j+1} / \|\hat{v}_{j+1}\|.$

     endFor
```

Notice that if $l \leq j$,

$$v_l^T \hat{v}_{j+1} = v_l^T A v_j - \sum_{i=1}^{j} h_{ij} \, v_l^T v_i$$

So, if $v_l^T v_i = \delta_{li}$ for $i, l \leq j$ then we have that $v_l^T \hat{v}_{j+1} = 0$. Hence, if $\hat{v}_{j+1} \neq 0$, we have that $v_l^T v_i = 0$ for $i, l \leq j+1$.

Next, let us show that if $\dim K_k = k$, then the algorithm (1), (2) does <u>not</u> break down.

Let us proceed by induction on k. For $k=1$, the result is obvious. Assume the result holds for $k=m$ and let us show it holds for $k=m+1$. Since it holds for $k=m$, we have that

$$K_m = \operatorname{span}\{A^i r_0\}_{i=0}^{m-1}$$

$$= \operatorname{span}\{A^{i-1} v_1\}_{i=1}^{m}$$

$$= \operatorname{span}\{v_1, \dots, v_m\}.$$

Now, this implies that

$$K_{m+1} = \operatorname{span}\{v_1, \dots, v_m, A^m v_1\}.$$

By (3) and (4), for $j \le m$,

$$\hat{v}_{j+1} = A v_j - \sum_{i=1}^{j} h_{ij} v_i$$

$$\Rightarrow \quad A v_j = \sum_{i=1}^{j+1} h_{ij} v_i$$

$$\Rightarrow \quad A v_j \in \operatorname{span}\{v_1, \dots, v_{j+1}\}$$
$$\Rightarrow \quad A^m v_1 \in \operatorname{span}\{v_1, \dots, v_{m+1}\}.$$

and so, $v_{m+1} \neq 0$. This implies that the algorithm does not break down.

Finally, notice that if $m > j+1$,

$$v_m^T A v_j = \sum_{i=1}^{j+1} h_{ij}\, v_m^T v_i \quad = 0,$$

by the inductive hypothesis. This shows that H_k is an upper Hessember matrix which is tridiagonal if A is symmetric.

③ The Full Orthogonalization Method (FOM)

The above discussion motivates the introduction of the FOM:

- Pick the initial guess x_0, compute $r_0 = Ax_0 - b$ and set $v_1 = r_0 / \|r_0\|$
- For $j = 1, 2, \ldots, k$ do

```
For $j = 1, 2, \ldots, k$ do
    $h_{ij} = v_i^T A v_j \quad i = 1, 2, \ldots, j$
    $\hat{v}_{j+1} = A v_j - \sum_{i=1}^{j} h_{ij} v_i$
    $h_{j+1\,j} = \|\hat{v}_{j+1}\|$
    $v_{j+1} = \hat{v}_{j+1} / h_{j+1,j}$
endFor
```

- Set $x_k = x_0 + V_k y_k$ where

$$y_k = -H_k^{-1} \|r_0\|\, e_1$$

Note that the algorithm breaksdown only if for some $j \leq k$, $h_{j+1\,j} = 0$. Next we show that this happens only if $r^j = 0$!

The relations (3) can be rewritten in matrix form as

$$A V_k = V_{k+1} \bar{H}_k .$$

Note that

$$H_k = V_k^T A V_k = V_k^T V_{k+1} \bar{H}_k ,$$

hence

$$\bar{H}_k = \begin{bmatrix} H_k \\ 0 \, 0 \ldots 0 \; h_{k+1,k} \end{bmatrix} .$$

Since

$$x_k = x_0 + V_k y_k ,$$

we get

$$\begin{aligned} r_k &= r_0 + A V_k y_k \\ &= r_0 + V_{k+1} \bar{H}_k \, y_k \\ &= r_0 + V_k H_k \, y_k + h_{k+1,k} \, (e_k^T y_k) \, v_{k+1} \end{aligned}$$

$$\Rightarrow \quad r_k = h_{k+1,k}\,(e_k^T y_k)\, v_{k+1}.$$

this means that

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

and $r_k = 0$ if $h_{k+1,k} = 0$, as desired.

④ The GMRES method.

The difference, and only difference, between the FOM and the GMRES methods is the way in which y_k is computed. In FOM, y_k is computed by requesting that

$$\begin{aligned} r^k &= A(x_0 + V_k y_k) - b \\ &= r_0 + AV_k y_k \end{aligned}$$

be orthogonal to K_k. In GMRES, instead, it is asked that

$$\|r^k\|^2 = \|r^k(y_k)\|^2 =: J(y_k).$$

is a minimum, hence the name of the method!

Next, notice that

$$\begin{aligned} J(y) &= \| r_0 + A V_k y \|^2 \\ &= \| r_0 + V_{k+1} \bar{H}_k y \|^2 \\ &= \| \, \|r_0\| e_1 + \bar{H}_k y \|^2 \end{aligned}$$

This implies that

$$y_k = - (\bar{H}_k^T \bar{H}_k)^{-1} \bar{H}_k e_1 \quad \|r_0\| ,$$

and shows that this method is different than the FOM.

To actually compute y_k, we can still exploit the structure of J. Suppose that we obtain a $(k+1) \times (k+1)$ matrix Q_k, the accumulated product of rotation matrices, and an upper triangular matrix R_k such that

$$Q_k \bar{H}_k = R_k$$

(Note that the last row of R_k is zero!). Then

$$\begin{aligned} J(y) &= \| Q_k (\|r_0\| e_1 + \bar{H}_k y) \|^2 \\ &= \| \underbrace{\|r_0\| Q_k e_1}_{-g_k} + R_k y \|^2 \end{aligned}$$

Now if we write

$$g_k = \begin{bmatrix} \hat{g}_k \\ g_{k,k+1} \end{bmatrix}, \quad R_k = \begin{bmatrix} \hat{R}_k \\ 0 \dots 0 \end{bmatrix},$$

then $\quad y_k = \hat{R}_k^{-1} \hat{g}_k \quad$ and $\quad J(y_k) = |g_{k,k+1}|$.

5 Convergence analysis

To analyze the GMRES method, we begin by noting that if the Arnoldi process does not break down, then $\hat{R}_k$ is invertible. To see this, consider the beginning of the algorithm leading to the equation $\varphi_k \bar{H}_k = R_k$:

$$\left[\begin{array}{cc|ccc} c & -s & 0 & 0 & 0 \\ s & c & 0 & 0 & 0 \\ \hline 0 & 0 & & & \\ 0 & 0 & & \text{Id} & \\ 0 & 0 & & & \end{array}\right] \begin{bmatrix} r & x & x & x \\ h & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \\ 0 & 0 & 0 & x \end{bmatrix} = \begin{bmatrix} t & y & y & y \\ 0 & y & y & y \\ 0 & x & x & x \\ 0 & 0 & x & x \\ 0 & 0 & 0 & x \end{bmatrix}$$

where $\quad c = \dfrac{r}{\sqrt{r^2+h^2}}, \quad s = -\dfrac{h}{\sqrt{r^2+h^2}}, \quad t = \sqrt{r^2+h^2}$

We see that if $h \neq 0$, then $t > 0$. Now notice that in every multiplication by a rotation, $h = h_{j+1,j}$, which is non-zero because the Arnoldi process did not break down for $j \leq k$! This establishes that $\hat{R}_k$ is invertible.

this means that the GMRES method can only break down if for some $j < N$, $h_{j+1\,j} = 0$. Next, we prove that this happens if and only if $r_j = 0$, that is, if and only if $Ax_j = b$.

Assume that $h_{j+1\,j} = 0$. this implies that

$$AV_j = V_j H_j .$$

In this case, we have

$$\begin{aligned} J(y) &= \| r_0 + AV_j\, y \|^2 \\ &= \| r_0 + V_j H_j\, y \|^2 \\ &= \| V_j\, (\|r_0\| e_1 + H_j y) \|^2 \\ &= \| \|r_0\| e_1 + H_j y \|^2 \\ &= 0 \end{aligned}$$

if

$$y = - \|r_0\| H_j^{-1} e_1 .$$

this means that $r_j = 0$, as claimed.

Assume now that $r_j = 0$ and that $r_i \neq 0$ $i < j$. then

$$\|r_j\| = | \overset{(-\|r_0\| Q_{j+1} e_1)_{j+1}}{\overset{\parallel}{g_{j,j+1}}} | = |s_j| \, | \overset{(-\|r_0\| Q_j e_1)_j}{\overset{\parallel}{g_{j-1,j}}} | = |s_j| \, \|r_{j-1}\|$$

this implies $0 = s_j \Rightarrow h_{j+1,j} = 0$, as claimed.

We have thus proven the following result.

Theorem. The GMRES method converges in at most N iterations, where N = order of A.

We can also obtain an error estimate of the residual in some cases.

Theorem. Let A be a diagonalizable matrix. Then

$$\|r_m\|_2 \le K \, \epsilon_m \, \|r_0\|_2,$$

where

$$K = \|X\|_2 \, \|X^{-1}\|_2 ,$$
$$\epsilon_m = \min_{\substack{p\in P^m \\ p(0)=1}} \max_{1\le i\le N} |p(\lambda_i)|$$

and $\mathrm{diag}\{\lambda_i, i=1,\dots,N) = \Lambda = X^{-1} A X$.

Proof. Since

$$x_m = x_0 + V_m y_m ,$$

we have

$$r_m = r_0 + A V_m y_m$$

and since $V_m = K_m$,

$$r_m = p(A) r_0 \quad , \quad \text{degree } p = m, \quad p(0) = 1.$$

Since x_m <u>minimizes the residual over $x_0 + K_m$</u>, we get

$$\| r_m \|_2 = \min_{\substack{p \in P^m \\ p(0)=1}} \| p(A) r_0 \| .$$

But

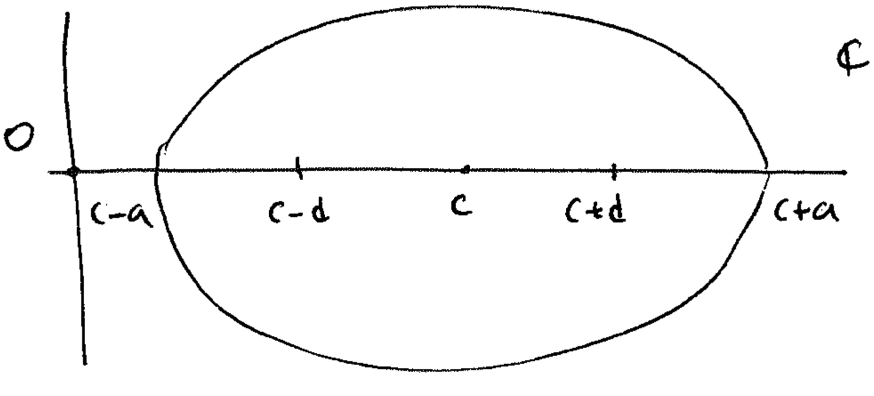
$$p(A) = X \; p(\Lambda) X^{-1}$$

and so

$$\| p(A) \|_2 \leq K \max_{1 \leq i \leq N} | p(\lambda_i) | .$$

This completes the proof. □

If all the eigenvalues of A are inside the ellipse of center $(c,0)$ focal distance d and major semi-axis a



$\mathbb{C}$
0
c−a
c−d
c
c+d
c+a