SGMRES and RB-SGMRES revisited

Gérard MEURANT

MTU60, Prague, February 2019

1. Introduction
2. SGMRES
3. RB-SGMRES
4. Numerical experiments
5. Conclusions

GMRES is one of the most used methods for solving nonsymmetric linear systems $Ax = b$

It uses an orthonormal basis of the Krylov subspace

$$\mathcal{K}_k(A, r_0) = \{r_0, Ar_0, A^2 r_0, \ldots, A^{k-1} r_0\}$$

constructed with the Arnoldi process and $r_0 = b - Ax_0$

$$AV_{n,k} = V_{n,k+1} \underline{H}_k$$

where $\underline{H}_k$ is a $(k+1) \times k$ upper Hessenberg matrix and $V_{n,k}^* V_{n,k} = I_k$

The iterates are defined as

$$x_k = x_0 + V_{n,k}\, y_k$$

y_k is computed by minimizing the residual norm

$$\min_y \| \, \|r_0\| e_1 - \underline{H}_k y \|$$

This is done by using Givens rotations to reduce $\underline{H}_k$ to upper triangular form and solving a triangular (small) linear system

Simpler GMRES

H.F. Walker and L. Zhou, *A simpler GMRES*, Numer. Linear Algebra Appl., v 1 n 6 (1994), pp. 571–581

Let us assume that we have a basis of $\mathcal{K}_k(A, b)$ given by the columns of a matrix $V_{n,k}$

This basis is not necessarily orthogonal but the basis vectors are of unit norm

Let $W_{n,k}$ be an orthonormal matrix obtained by the QR factorization of $AV_{n,k}$

$$AV_{n,k} = W_{n,k} T_k$$

where T_k is upper triangular

The columns of $W_{n,k}$ give an orthonormal basis of $A\mathcal{K}_k(A, r_0)$

To minimize the residual norm we would like to have residual vectors which are orthogonal to $A\mathcal{K}_k(A, r_0)$ that is,

$$r_k = (I - W_{n,k} W_{n,k}^*) r_0$$

This is obtained recursively by

$$r_k = r_{k-1} - \alpha_k w_k, \quad \alpha_k = (w_k, r_{k-1})$$

As in GMRES we define the iterates as

$$x_k = x_0 + V_{n,k}\, y_k$$

It yields

$$r_k = r_0 - W_{n,k}\, T_k\, y_k$$

Multiplying by $W_{n,k}^*$,

$$T_k y_k = W_{n,k}^* r_0 = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}$$

since $(w_j, r_0) = (w_j, r_{j-1}) = \alpha_j$

We have to choose $V_{n,k}$. In SGMRES Walker and Zhou chose

$$V_{n,k} = \begin{pmatrix} \frac{r_0}{\|r_0\|} & W_{n,k-1} \end{pmatrix}$$

They proved that if r_0 is not in $A\mathcal{K}_{k-1}(A, r_0)$, the columns of $V_{n,k}$ form a basis of $\mathcal{K}_k(A, r_0)$ and

$$\frac{\|r_0\|}{\|r_{k-1}\|} \leq \mathcal{K}(V_{n,k}) \leq 2\frac{\|r_0\|}{\|r_{k-1}\|}$$

Theorem

Let $\alpha^{(k-1)} = \begin{pmatrix} \alpha_1 & \cdots & \alpha_{k-1} \end{pmatrix}^T$. *Then,*

$$\mathcal{K}(V_{n,k}) = \frac{\|r_0\| + \|\alpha^{(k-1)}\|}{\|r_{k-1}\|}.$$

Proof. Let $\tilde{r}_0 = r_0/\|r_0\|$. Then,

$$V_{n,k} = \begin{pmatrix} \tilde{r}_0 & W_{n,k-1} \end{pmatrix}$$

and, since $W_{n,k}^* W_{n,k} = I$,

$$V_{n,k}^* V_{n,k} = \begin{pmatrix} 1 & \tilde{r}_0^* W_{n,k-1} \\ W_{n,k-1}^* \tilde{r}_0 & I \end{pmatrix}$$

But, $W_{n,k-1}^* r_0 = \alpha^{(k-1)}$. It yields

$$V_{n,k}^* V_{n,k} = I - \frac{1}{\|r_0\|} \begin{pmatrix} 0 & [\alpha^{(k-1)}]^* \\ \alpha^{(k-1)} & 0 \end{pmatrix} = I - \frac{1}{\|r_0\|} B_k$$

The two nonzero eigenvalues of B_k are $\pm\|\alpha^{(k-1)}\|$

Therefore, the singular values of $V_{n,k}$ are $k-2$ times 1 and $\sqrt{1 \pm \frac{\|\alpha^{(k-1)}\|}{\|r_0\|}}$ and its condition number is

$$\mathcal{K}(V_{n,k}) = \left(\frac{\|r_0\| + \|\alpha^{(k-1)}\|}{\|r_0\| - \|\alpha^{(k-1)}\|}\right)^{\frac{1}{2}} = \frac{\|r_0\| + \|\alpha^{(k-1)}\|}{(\|r_0\|^2 - \|\alpha^{(k-1)}\|^2)^{\frac{1}{2}}}$$

But, $r_{k-1} = r_0 - W_{n,k-1}\alpha^{(k-1)}$ and, because of the orthogonality properties,

$$\|r_{k-1}\|^2 = \|r_0\|^2 - \|W_{n,k_1}\alpha^{(k-1)}\|^2 = \|r_0\|^2 - \|\alpha^{(k-1)}\|^2$$

□

When the residual norms go to zero, $V_{n,k}$ increases unboundedly

The interest of SGMRES compared to GMRES is that we do not have to compute the QR factorization of the upper Hessenberg matrix $\underline{H}_k$ using Givens rotations

RB-SGMRES

P. JIRÁNEK, M. ROZLOŽNÍK AND M.H. GUTKNECHT, *How to make simpler GMRES and GCR more stable*, SIAM J. Matrix Anal. Appl., v 30 n 4 (2008), pp. 1483–1499

They chose

$$V_{n,k} = \begin{pmatrix} \frac{r_0}{\|r_0\|} & \cdots & \frac{r_{k-1}}{\|r_{k-1}\|} \end{pmatrix}$$

RB means residual-based

If r_0 is not in $A\mathcal{K}_{k-1}(A, r_0)$ and the residual norms are strictly decreasing, the columns of $V_{n,k}$ are linearly independent
They proved that

$$1 \leq \mathcal{K}(V_{n,k}) \leq \sqrt{k}\gamma_k, \quad \gamma_k = \left(1 + \sum_{j=1}^{k-1} \frac{\|r_{j-1}\|^2 + \|r_j\|^2}{\|r_{j-1}\|^2 - \|r_j\|^2}\right)^{\frac{1}{2}}$$

It turns out that we can explicitly compute the entries of $S_k = V_{n,k}^* V_{n,k}$

Proposition

The entries of S_k *are given by*

$$[S_k]_{i,i} = 1, \quad [S_k]_{i,j} = \frac{\|r_i\|}{\|r_j\|}, \ i > j, \quad i = 1, \dots, k$$

Proof. Let $\tilde{r}_i = r_i / \|r_i\|$. By definition the entry (i,j) of S_k is $(\tilde{r}_i, \tilde{r}_j)$.
It is obvious that $[S_k]_{i,i} = 1$, $i = 1, \dots, k$. Then,

$$\begin{aligned}
(r_i, r_j) &= (r_0 - W_{n,i}\alpha^{(i)}, r_0 - W_{n,j}\alpha^{(j)}), \\
&= \|r_0\|^2 + (\alpha^{(i)}, W_{n,i}^* W_{n,j}\alpha^{(j)}) - (r_0, W_{n,j}\alpha^{(j)}) - (W_{n,i}\alpha^{(i)}, r_0), \\
&= \|r_0\|^2 + (\alpha^{(i)}, W_{n,i}^* W_{n,j}\alpha^{(j)}) - \|\alpha^{(j)}\|^2 - \|\alpha^{(i)}\|^2
\end{aligned}$$

If $i > j$,

$$W_{n,i}^* W_{n,j} \alpha^{(j)} = \begin{pmatrix} \alpha^{(j)} \\ 0 \end{pmatrix}$$

Hence $(\alpha^{(i)}, W_{n,i}^* W_{n,j} \alpha^{(j)}) = \|\alpha^{(j)}\|^2$ and

$$(r_i, r_j) = \|r_0\|^2 - \|\alpha^{(i)}\|^2$$

We note that $\|r_i\|^2 = \|r_0\|^2 - \|\alpha^{(i)}\|^2$ and, dividing by $\|r_i\| \, \|r_j\|$ we obtain

$$(\tilde{r}_i, \tilde{r}_j) = \frac{\|r_0\|^2 - \|\alpha^{(i)}\|^2}{\|r_i\| \, \|r_j\|} = \frac{\|r_i\|^2}{\|r_i\| \, \|r_j\|} = \frac{\|r_i\|}{\|r_j\|}$$

which proves the result □

$$V_{n,k}^* V_{n,k} = S_k = \begin{pmatrix} 1 & \frac{\|r_1\|}{\|r_0\|} & \frac{\|r_2\|}{\|r_0\|} & \cdots & \frac{\|r_{k-1}\|}{\|r_0\|} \\ \frac{\|r_1\|}{\|r_0\|} & 1 & \frac{\|r_2\|}{\|r_1\|} & \cdots & \frac{\|r_{k-1}\|}{\|r_1\|} \\ \frac{\|r_2\|}{\|r_0\|} & \frac{\|r_2\|}{\|r_1\|} & 1 & \cdots & \frac{\|r_{k-1}\|}{\|r_2\|} \\ \vdots & \vdots & & \ddots & \vdots \\ \frac{\|r_{k-1}\|}{\|r_0\|} & \frac{\|r_{k-1}\|}{\|r_1\|} & \cdots & & 1 \end{pmatrix}$$

This shows that the inverse of $V_{n,k}^* V_{n,k}$ is a symmetric tridiagonal matrix $(\beta_{i-1}, \alpha_i, \beta_i)$ whose nonzero entries ar not too difficult to compute

Proposition

The non zero coefficients of the inverse of the matrix $V_{n,k}^* V_{n,k}$ *are given by*

$$\begin{aligned}
\alpha_1 &= \frac{\|r_0\|^2}{\|r_0\|^2 - \|r_1\|^2}, \quad \beta_1 = -\frac{\|r_0\| \, \|r_1\|}{\|r_0\|^2 - \|r_1\|^2}, \\
\alpha_i &= \|r_{i-1}\|^2 \left(\frac{1}{\|r_{i-2}\|^2 - \|r_{i-1}\|^2} + \frac{1}{\|r_{i-1}\|^2 - \|r_i\|^2} \right), \\
& \qquad \text{for } i = 2, \ldots, k-1, \\
\beta_i &= -\frac{\|r_{i-1}\| \, \|r_i\|}{\|r_{i-1}\|^2 - \|r_i\|^2}, \; i = 2, \ldots, k-1, \\
\alpha_k &= \frac{\|r_{k-2}\|^2}{\|r_{k-2}\|^2 - \|r_{k-1}\|^2}
\end{aligned}$$

This proposition can be used to bound the smallest singular value of the matrix $V_{n,k}$ as well as its condition number

Theorem

Let

$$\gamma_1 = \frac{\|r_0\|}{\|r_0\| - \|r_1\|},$$

$$\gamma_i = \|r_{i-1}\| \left(\frac{1}{\|r_{i-2}\| - \|r_{i-1}\|} + \frac{1}{\|r_{i-1}\| - \|r_i\|} \right), \quad i = 2, \ldots, k-1,$$

$$\gamma_k = \frac{\|r_{k-2}\|}{\|r_{k-2}\| - \|r_{k-1}\|}$$

Then,

$$\sigma_{\min}(V_{n,k}) \geq \frac{1}{\max_{i=1,\ldots,k} \sqrt{\gamma_i}}$$

and the condition number of $V_{n,k}$ *is bounded by*

$$\mathcal{K}(V_{n,k}) \leq \sqrt{k} \max_{i=1,\ldots,k} \sqrt{\gamma_i}$$

Proof. We use the previous proposition and the Gershgorin intervals to bound the largest eigenvalue of $V_{n,k}^* V_{n,k}$

Taking the square root, this gives a lower bound for the smallest singular value of $V_{n,k}$

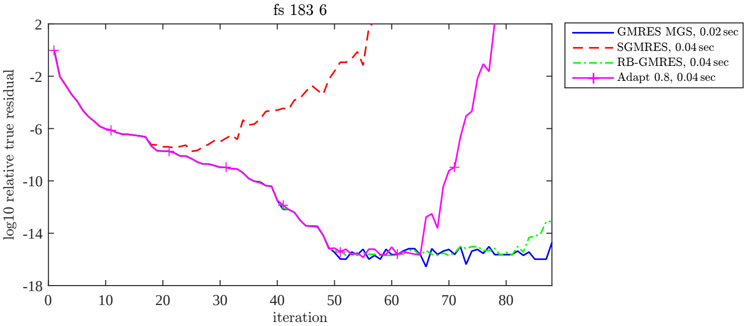
Since the columns of $V_{n,k}$ are of unit norm, its Frobenius norm is bounded by $\sqrt{k}$ which yields the upper bound for the condition number □

We can sometimes improve the upper bound $\sqrt{k}$ for $\|V_{n,k}\|$ by considering the lower bound obtained for the Gershgorin intervals (if it is positive)

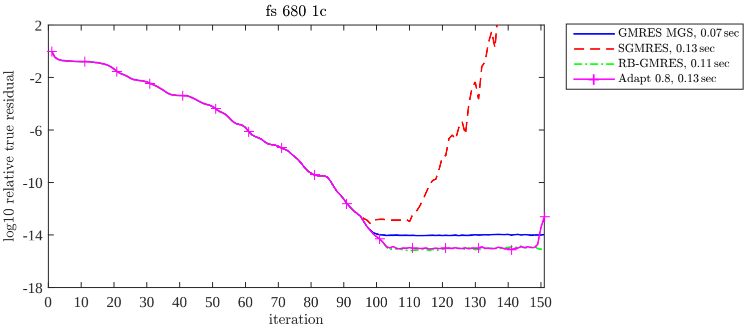
An adaptive version mixing SGMRES and RB-SGMRES was proposed in

P. JIRÁNEK AND M. ROZLOŽNÍK, *Adaptive version of Simpler GMRES*, Numer. Algorithms v 53 n 1 (2010), pp. 93–112.

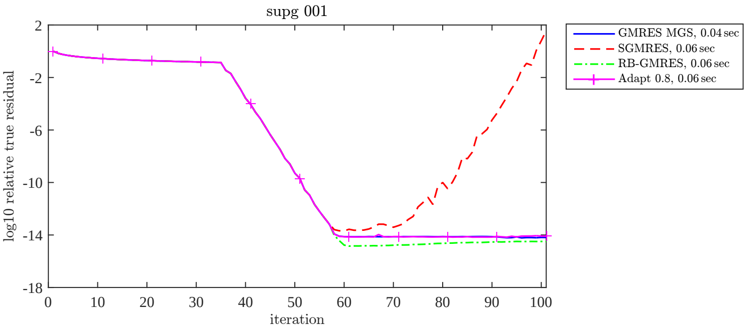
Numerical experiments



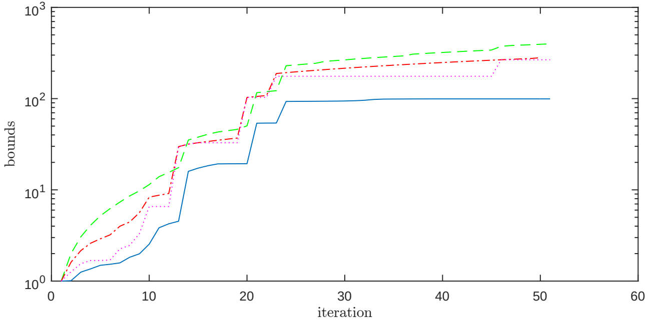
fs 183 6, relative true residual norms, GMRES-MGS (plain), SGMRES (dashed), RB GMRES (dot-dashed), ASGMRES $\nu = 0.8$ $(+)$



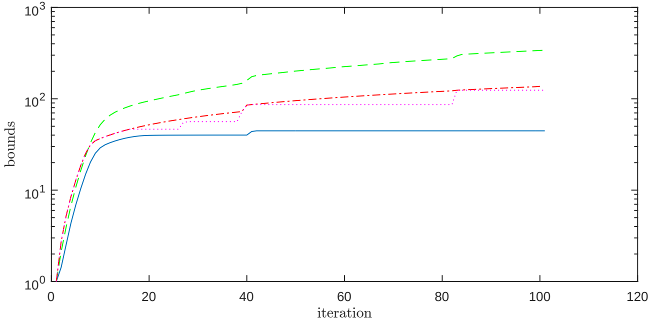
fs 680 1c, relative true residual norms, GMRES-MGS (plain), SGMRES (dashed), RB GMRES (dot-dashed), ASGMRES $\nu = 0.8$ $(+)$



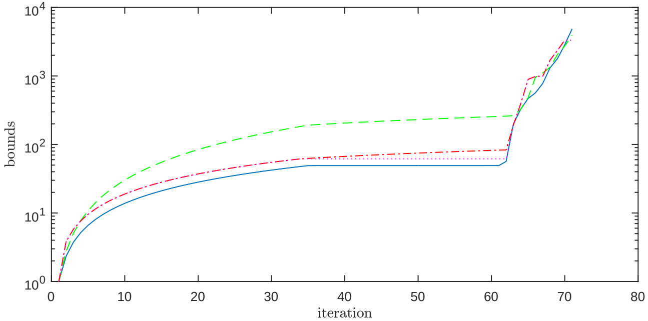
supg 001 1225, relative true residual norms, GMRES-MGS (plain), SGMRES (dashed), RB GMRES (dot-dashed), ASGMRES $\nu = 0.8$ (+)



fs 183 6, RB-SGMRES, bounds for $\mathcal{K}(V_{n,k})$, JRG (green), GM 1 (red), GM 2 (magenta)



fs 680 1c, RB-SGMRES, bounds for $\mathcal{K}(V_{n,k})$, JRG (green), GM 1 (red), GM 2 (magenta)



supg 001 1225, RB-SGMRES, bounds for $\mathcal{K}(V_{n,k})$, JRG (green), GM 1 (red), GM 2 (magenta)

Conclusions

As it is known SGMRES is quite unstable

We gave an exact expression for the condition number of $V_{n,k}$ in SGMRES

RB-SGMRES is a much better method which suffers problems only when GMRES stagnates

We gave a new bound on the condition number of $V_{n,k}$ in RB-SGMRES

Happy birthday, Mirek