Rank one updates that accelerate restarted GMRES

J. Duintjer Tebbens

Institute of Computer Science, Academy of Science of the Czech Republic, Prague

1 Introduction

The GMRES method [14] ranks among the most popular methods to solve a large, sparse and nonsymmetric linear system

$$\mathbf{A}x = b, \tag{1}$$

where the matrix $\mathbf{A} \in \mathrm{R}^{n\times n}$ is nonsingular and $b, x \in \mathrm{R}^n$. The kth GMRES iterate x_k minimizes the residual norm $\|b - \mathbf{A}s\|$ over all elements s of the affine subspace

$$x_0 + \mathcal{K}_k(\mathbf{A}, r_0), \tag{2}$$

where x_0 is the initial guess, $r_0 = b - \mathbf{A}x_0$ and $\mathcal{K}_k(\mathbf{A}, r_0)$ is the kth Krylov subspace,

$$\mathcal{K}_k(\mathbf{A}, r_0) \equiv \mathrm{span}\{r_0, \mathbf{A}r_0, \ldots, \mathbf{A}^{k-1}r_0\}.$$

To compute iterates, the GMRES method exploits an orthogonal basis $\mathbf{V}_k$ of $\mathcal{K}_k(\mathbf{A}, r_0)$ satisfying

$$\mathbf{A}\mathbf{V}_k = \mathbf{V}_k\mathbf{H}_k + \tilde{v}_{k+1}e_k^T, \tag{3}$$

where $\mathbf{H}_k \in \mathrm{R}^{k\times k}$ is upper Hessenberg and $\tilde{v}_{k+1}$ is the unscaled $(k+1)$st basis vector. Because of the residual minimizing property, GMRES convergence curves do not increase and, in exact arithmetics, the solution is found after at most n steps. Although in practice we are often interested only in a good approximation of the exact solution, in many applications the system dimension is so large that even the computation of a satisfactory approximation is too expensive for today's high performance computers. This is due to the fact that in the GMRES method computational and storage costs per iteration grow with the iteration number. A currently used remedy is restarting the method after a fixed, relatively small number of steps. As GMRES residual norms do not increase, every restart brings us closer to the solution, or, in the worst case, it leaves the approximation unchanged. The latter scenario is called *stagnation* and represents the main drawback of restarted GMRES. Many techniques to overcome stagnation have been proposed in the literature, for example augmentation of Krylov subspaces [9, 10], implicit polynomial filtration [12] of Arnoldi vectors, deflation through preconditioning [2, 5], exploitation of inner-outer cycles [13, 15]. In this contribution we describe an alternative strategy to accelerate restarted GMRES. It is based on a specific rank one update of the system matrix [3, 4].

2 A rank one updated system matrix

Consider the modified system matrix

$$\hat{\mathbf{A}} := \mathbf{A} - by^T, \tag{4}$$

where b is the right hand side of (1) and $y \in \mathrm{R}^n$ is a free parameter vector. If we assume $\hat{\mathbf{A}}$ is nonsingular, then with the Sherman-Morrison formula (see e.g. [8]), the solution x of (1) equals

$$x = \mathbf{A}^{-1}b = (\hat{\mathbf{A}} + by^T)^{-1}b = \hat{\mathbf{A}}^{-1}b - \frac{y^T\hat{\mathbf{A}}^{-1}b}{1 + y^T\hat{\mathbf{A}}^{-1}b}\,\hat{\mathbf{A}}^{-1}b. \tag{5}$$

Computation of the right hand side of (5) involves solving the *auxiliary system* defined as

$$\hat{\mathbf{A}}\hat{x} = b. \tag{6}$$

Once a satisfactory approximation $\hat{x}_k$ to the solution $\hat{x}$ of (6) is found, we obtain the *back-transformed iterate* $\bar{x}_k$ for the original system (1) by approximating (5) as

$$x = \mathbf{A}^{-1}b = \hat{x} - \frac{y^T\hat{x}}{1 + y^T\hat{x}}\hat{x} = \frac{1}{1 + y^T\hat{x}}\hat{x} \approx \frac{1}{1 + y^T\hat{x}_k}\hat{x}_k \equiv \bar{x}_k \in \hat{x}_0 + \mathcal{K}_k(\hat{\mathbf{A}}, \hat{r}_0), \tag{7}$$

provided $y^T\hat{x}_k \neq -1$. Solving (1) with formula (7) makes sense when the residual norms of the auxiliary system (6) converge faster than those of the original system. This is the case when the Krylov subspaces of the auxiliary system have more favorable properties with regards to the GMRES process. The kth Krylov subspace generated by $\hat{\mathbf{A}}$ equals

$$\mathcal{K}_k(\hat{\mathbf{A}}, \hat{r}_0) = \mathcal{K}_k(\mathbf{A} - by^T, \hat{r}_0) = \mathrm{span}\{\hat{r}_0, \mathbf{A}\hat{r}_0 - (y^T\hat{r}_0)\,b\,, \ldots\,, (\mathbf{A} - by^T)^{k-1}\hat{r}_0\}.$$

There are good reasons for using the zero initial guess in GMRES processes [11]. With the initial guesses $\hat{x}_0 = 0 = x_0$,

$$\mathcal{K}_k(\hat{\mathbf{A}}, \hat{r}_0) = \mathcal{K}_k(\hat{\mathbf{A}}, b) = \mathcal{K}_k(\mathbf{A}, b) = \mathcal{K}_k(\mathbf{A}, r_0), \tag{8}$$

and consequently, iterates of the form (2) for (1) are taken from the same subspace as the back-transformed iterates (7). As GMRES iterates are defined by their residual norm minimizing property, this means that $\bar{x}_k$ can never yield a smaller residual norm than x_k. But with nonzero initial guesses back-transformed iterates might yield residual norms that converge faster than the ones obtained from the original iterates of the form (2).

When we apply the *restarted* GMRES method, then the initial guess of every restart cycle consists of the approximation obtained from the previous cycle and is in general not zero anymore. Hence the Krylov subspaces generated by (6) can differ from those belonging to (1) and we may attempt to improve the spaces of (6) through special choices of the parameter vector y. This is the main idea of our approach. Having found such an y, we restart GMRES applied to (6) until we have found a satisfactory approximation $\hat{x}_k$ of $\hat{x}$ and we back-transform according to (7).

3 Convergence of the rank one updated system

We now concentrate on the choice of the parameter vector y in (4). A good choice will yield an auxiliary system (6) whose residual norms converge faster than those of the original system when restarted GMRES is applied. In addition, back-transformation with (7) should not cause a serious loss of quality. We briefly discuss two classes of choices of y, connected with two different classes of system matrices.

If the matrix $\mathbf{A}$ is normal, convergence can be related to eigenvalues and we may design a parameter vector that eliminates convergence hampering eigenvalues. The following result has been proved in [4].

Theorem 3.1 *Let the matrix $\mathbf{B} \in R^{l \times l}$ and the vector $c \in R^l$ be such that the Krylov subspace $\mathcal{K}_l(\mathbf{B}, c)$ has full dimension and let $\{\theta_1, \ldots, \theta_l\}$ be a set of real and complex conjugate values. Then there exists a vector $z \in R^l$ such that $\mathbf{B} - cz^T$ has the eigenvalues $\theta_1, \ldots, \theta_l$.*

The proof, which was inspired by the proof of the main result in [7], shows that prescription of the spectrum of the modified system matrix (4) is too expensive for a large system dimension n. The eigenvalues of the small Hessenberg matrix $\mathbf{H}_k = \mathbf{V}_k^T \mathbf{A} \mathbf{V}_k$ from (3), however, approximate the eigenvalues of $\mathbf{A}$. Moreover, the spectrum of the Hessenberg matrix $\hat{\mathbf{H}}_k = \mathbf{V}_k^T \hat{\mathbf{A}} \mathbf{V}_k = \mathbf{H}_k - \mathbf{V}_k^T b y^T \mathbf{V}_k$ generated by GMRES applied to (6) with $\hat{x}_0 = 0$, approximates the spectrum of $\hat{\mathbf{A}}$. Hence with Theorem 3.1 we can prescribe the eigenvalues of $\hat{\mathbf{H}}_k$, and presume that the eigenvalues of $\hat{\mathbf{A}}$ are close. Using this strategy, one has to assure that the rank one matrix by^T does not remove $\mathbf{A}$ too far from normality, otherwise it is not certain anymore whether a ,,harmless" spectrum has any positive influence on convergence [7].

In the nonnormal case, it is in general not clear what properties of $\mathbf{A}$ and r_0 cause restarted GMRES to converge slowly or even to stagnate. It has been proved that given a non-increasing sequence of real values and a set of real and complex conjugate values, there exists a class of right hand sides and matrices whose spectrum is the set of real and complex conjugate values such that GMRES applied to linear systems with these matrices and right hand sides yields a sequence of residual norms equal to the given sequence of real values [1, 6, 7]. In the special case of a system matrix of the form $\hat{\mathbf{A}} = \mathbf{A} - by^T$ we can choose y such that $\hat{\mathbf{A}}$ belongs to the class of matrices generating a given convergence speed [3, 4]:

Theorem 3.2 *Let $f_0 \geq f_1 \geq f_2 \ldots \geq f_k > 0$, $k < n$, be a non-increasing sequence of real values with $\|b\| = f_0$ and let us choose the initial guess $\hat{x}_0 = 0$ to solve the system (6) with system matrix $\hat{\mathbf{A}} = \mathbf{A} - by^T$. If the Krylov subspace $\mathcal{K}_k(\mathbf{A}, b)$ has full dimension, then there exists at least one $y \in R^n$ such that the residual vectors $\hat{r}_j$ obtained by application of the GMRES method to the system (6) satisfy*

$$\|\hat{r}_j\| = f_j, \quad 0 \leq j \leq k.$$

The proof is based on the observation that the convergence speed of GMRES depends on the subspaces $\hat{\mathbf{A}}\mathcal{K}_j(\hat{\mathbf{A}}, b) = (\mathbf{A} - by^T)\mathcal{K}_j(\mathbf{A} - by^T, b)$, see [6], which we can modify with the parameter vector y. We propose to apply Theorem 3.2 by defining an auxiliary system whose first k residual norms converge faster than the corresponding norms of the original system. When restarting this auxiliary system we might expect that also the first k iterations (or possibly even more) of every restart cycle converge fast and that back-transformation with (7) accelerates the original restarted process, or even overcomes its stagnation.

4 Open Questions

The techniques mentioned in the previous section have shown to be able to overcome stagnation of restarted GMRES in many numerical experiments. When we try to modify the spectrum of normal matrices, the succeeding of our approach depended essentially on the quality of the used approximate eigenvalues. This holds in fact for many other GMRES accelerating techniques too. One has to carefully compare these with our strategy to be able to say which of them is fastest, most stable, least demanding with regards to storage costs, etc...

When matrices are further from normal and when we use the more heuristical idea of prescribing initial residual norms, it is less clear why the one stagnation can be overcome but the next could

not. In particular, the influence of the choice of initial residual norms on the quality of the back-transformation has to be better understood. In addition, we were surprised to notice that the strategy worked especially well for many linear systems arising from discretized partial differential equations, but we are not able to say why this is so.

Acknowledgement: This work was supported by the Program Information Society under project 1ET400300415.

References

[1] M. Arioli, V. Pták and Z. Strakoš, *Krylov Sequences of maximal length and convergence of GMRES*, B.I.T., 38(1998), pp. 636–643.

[2] J. Baglama, D. Calvetti, G. Golub and L. Reichel, *Adaptively Preconditioned GMRES Algorithms*, SIAM J. Sci. Comput., 20(1998), pp. 243–269.

[3] E. J. Duintjer Tebbens, *An Application of the Sherman-Morrison Formula to the GMRES Method*, in Conjugate Gradients Algorithms and Finite Elements Methods (Jyväskylä, 2002), Springer-Verlag Berlin(2004), pp. 69–92.

[4] E. J. Duintjer Tebbens, *Modern Methods for Solving Linear Systems*, PhD Thesis, Charles University Prague (2004).

[5] K. Burrage, J. Erhel, B. Pohl, *Restarted GMRES Preconditioned by Deflation*, Elsevier J. Comput. Appl. Math., 69(1996), pp. 303–318.

[6] A. Greenbaum and Z. Strakoš, *Matrices that generate the same Krylov Residual Spaces*, in Recent Advances in Iterative Methods, G.H. Golub, A. Greenbaum and M. Luskin, eds, 50(1994), pp. 95–118.

[7] A. Greenbaum, V. Pták and Z. Strakoš, *Any Nonincreasing Convergence Curve is Possible for GMRES*, SIAM J. Matrix Anal. Appl., 17(1996), pp. 465–469.

[8] W. W. Hager, *Updating the Inverse of a Matrix*, SIAM Review, 31(1989), pp. 221–239.

[9] R. B. Morgan, *A Restarted GMRES Method Augmented with Eigenvectors*, SIAM J. Matrix Anal. Appl., 16(1995), pp. 1154–1171.

[10] R. B. Morgan, *Implicitly Restarted GMRES and Arnoldi Methods for Nonsymmetric Systems of Equations*, SIAM J. Matrix Anal. Appl., 21(2000), pp. 1112–1135.

[11] C. C. Paige and Z. Strakoš, *Residual and Backward Error Bounds in Minimum Residual Krylov Subspace Methods*, SIAM J. Sci. Comput., 23(2002), pp. 1899–1924.

[12] D. C. Sorensen, *Implicit Application of Polynomial Filters in k-Step Arnoldi Method*, SIAM J. Matrix Anal. Appl., 13(1992), pp. 357–385.

[13] Y. Saad, *A Flexible Inner-Outer Preconditioned GMRES Algorithm*, SIAM J. Sci. Stat. Comput., 14(1993), pp. 461–469.

[14] Y. Saad, M. H. Schultz, *GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems*, SIAM J. Sci. Stat. Comput., 7(1986), pp. 856–869.

[15] H. A. Van der Vorst, C. Vuik, *GMRESR: A Family of Nested GMRES Methods*, Numer. Linear Algebra Applications, 1(1994), pp. 369–386.