Rank One Updates that Accelerate Restarted GMRES

Jurjen Duintjer Tebbens

Institute of Computer Science

Academy of Sciences of the Czech Republic

Seminar on Numerical Analysis (SNA'05), Ostrava, February 7-11, 2005

Outline

1. Introduction
2. The rank one updated system matrix
3. Acceleration with the rank one updated system
   (a) The normal case
   (b) The non-normal case
4. Open questions

1. Introduction

We consider a linear system

$$\mathbf{A}x = b$$

- $\mathbf{A}$ is $n \times n$, nonsingular and nonsymmetric
- $\mathbf{A}$ is large and sparse

We focuss on the GMRES method [Saad, Schultz - 86]: Starting with an initial guess x_0, iterates x_k minimize

$$\|b - \mathbf{A}s\| \quad \text{over all} \quad s \in x_0 + \mathcal{K}_k(\mathbf{A}, r_0),$$

where $r_0 = b - \mathbf{A}x_0$ and

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

Because of the residual minimizing property, GMRES convergence curves do not increase.

Here: Orthogonal basis of $\mathcal{K}_k(\mathbf{A}, r_0)$ given by columns of $\mathbf{V}_k$ with

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

where $\mathbf{H}_k \in \mathbb{R}^{k\times k}$ is upper Hessenberg and $\tilde{v}_{k+1}$ is the unscaled $(k+1)$st basis vector.

Computational and storage costs per iteration grow with the iteration number hence necessity to **restart**:

After every cycle of m steps, $m << n$, we restart the process with the initial guess x_m. Notation: GMRES(m).

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.

Techniques to overcome stagnation: Augmentation of Krylov subspaces [Morgan - 95], [Morgan - 00], deflation through preconditioning [Baglama et al - 98], [Burrage, Erhel, Pohl - 96], exploitation of inner-outer cycles [Saad - 93], [Van der Vorst, Vuik - 94].

Here we describe an alternative strategy to accelerate restarted GMRES. It is based on a specific rank one update of the system matrix.

2. The rank one updated system matrix

Consider the modified system matrix

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

where b is the right hand side of the system and $y \in \mathbb{R}^n$ is a free parameter vector.

Then

$$\mathbf{A}^{-1}b = (\widehat{\mathbf{A}} + by^T)^{-1}b = \widehat{\mathbf{A}}^{-1}b - \frac{y^T\widehat{\mathbf{A}}^{-1}b}{1 + y^T\widehat{\mathbf{A}}^{-1}b}\widehat{\mathbf{A}}^{-1}b.$$

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

$$\widehat{\mathbf{A}}\widehat{x} = b.$$

If GMRES(m) has found a satisfactory approximation $\widehat{x}_k$ to the solution $\widehat{x}$, we obtain the **back-transformed iterate** $\bar{x}_k$ for the original system by approximating as

$$\mathbf{A}^{-1}b = \widehat{x} - \frac{y^T\widehat{x}}{1 + y^T\widehat{x}}\widehat{x} = \frac{1}{1 + y^T\widehat{x}}\widehat{x} \approx \frac{1}{1 + y^T\widehat{x}_k}\widehat{x}_k \equiv \bar{x}_k\,,$$

provided $y^T \widehat{x}_k \neq -1$.

Solving with this formula makes sense when the residual norms for the auxiliary system converge faster than the residual norms for the original system. We will try to enforce this through special choices of the parameter vector y. This is the main idea of our approach.

2. Acceleration with the rank one updated system

The basic form of the algorithms we propose:

1. Apply m initial steps of GMRES to $\mathbf{A}x = b$ **with** $\mathbf{x_0 = 0}$ and gain information to construct the parameter vector y.

2. Apply GMRES(m) to $\widehat{\mathbf{A}}\widehat{x} = b$, **with $\mathbf{x_0 = 0}$ for the initial cycle**, and find a satisfactory approximation $\widehat{x}_k$.

3. Back-transform with $\frac{1}{1+y^T\widehat{x}_k}\widehat{x}_k = \bar{x}_k$.

The advantage of executing initial cycles with zero initial guesses is that with $x_0 = 0 = \widehat{x}_0$

$$\mathcal{K}_k(\widehat{\mathbf{A}}, b) = \text{span}\{b, (\mathbf{A} - by^T)b, \ldots, (\mathbf{A} - by^T)^{k-1}b\}.$$

Hence $\mathcal{K}_k(\widehat{\mathbf{A}}, b) = \mathcal{K}_k(\mathbf{A}, b)$ and information gained from the process applied to $\mathbf{A}x = b$ can be easily translated to a process for $\widehat{\mathbf{A}}\widehat{x} = b$.

In our quest for finding an $y \in \mathbb{R}^n$ such that GMRES(m) applied to

$$(\mathbf{A} - by^T)\widehat{x} = b$$

has fastest possible convergence speed, we were inspired by a series of articles [Greenbaum, Strakoš - 94], [Greenbaum, Pták, Strakoš - 96], [Arioli, Pták, Strakoš - 98] proving that

- Given a non-increasing convergence curve, a spectrum and a right hand side,

- there exist matrices with the given spectrum such that GM-RES applied to such a matrix and the given right hand side yields the given convergence curve.

Question: Can we construct an y such that $\mathbf{A} - by^T$ belongs to the class with the prescribed convergence curve and spectrum ?

Answer: We can either prescribe the spectrum or prescribe the convergence curve by the choice of y, but not both.

(a). The normal case

If the matrix **A** is normal, convergence can be related to eigenvalues and we may design a parameter vector that eliminates convergence hampering eigenvalues.

Theorem 1: *Let the matrix* $\mathbf{B} \in \mathbb{R}^{l \times l}$ *and the vector* $c \in \mathbb{R}^l$ *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* $z \in \mathbb{R}^l$ *such that* $\mathbf{B} - cz^T$ *has the eigenvalues* $\theta_1, \ldots, \theta_l$.

The proof, which was inspired by the proof in [Greenbaum, Pták, Strakoš - 96], shows that prescription of the eigenvalues of the matrix $\mathbf{A} - by^T$ is too expensive for a large system dimension n.

But: We have to our disposal a small matrix whose eigenvalues approximate the eigenvalues of $\mathbf{A}$:

$$\mathbf{H}_k = \mathbf{V}_k^T \mathbf{A} \mathbf{V}_k,$$

the restriction of $\mathbf{A}$ to $\mathcal{K}_k(\mathbf{A}, r_0)$.

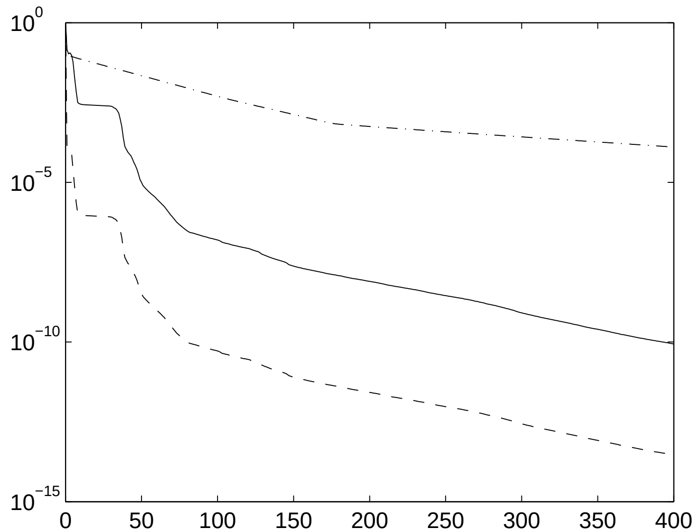
Moreover, with $\widehat{x}_0 = 0 = x_0$,

$$\widehat{\mathbf{H}}_k = \mathbf{V}_k^T(\mathbf{A} - by^T)\mathbf{V}_k = \mathbf{H}_k - \mathbf{V}_k^T by^T \mathbf{V}_k,$$

and the spectrum of $\widehat{\mathbf{H}}_k$ approximates the spec-trum of $\widehat{\mathbf{A}}$.

Hence with Theorem 1 we can prescribe the eigenvalues of $\widehat{\mathbf{H}}_k$, and presume that the eigenvalues of $\widehat{\mathbf{A}}$ are close.

Example: The shifted Laplacian $\mathbf{L} - 0.1\,\mathbf{I}$ of dimension 400. The smallest eigenvalues are -0.088 and -0.072. The Hessenberg matrix $\mathbf{H}_{15}$ has 2 negative eigenvalues -0.088 and -0.0452. We define an auxiliary Hessenberg matrix $\widehat{\mathbf{H}}_{15}$ where these smallest eigenvalues take the value 1. In the resulting auxiliary matrix -0.088 has vanished.



(b). The non-normal case

It is in general not clear what properties cause restarted GMRES to stagnate, hence here we will prescribe convergence curves instead of the spectrum.

Theorem 2: *Let* $\|b\| = f_0 \geq f_1 \geq f_2 \cdots \geq f_k > 0, k < n,$ *be a non-increasing sequence of real values. If* $\mathcal{K}_k(\mathbf{A}, b)$ *has full dimension, then there exists at least one* $y \in \mathbb{R}^n$ *such that the residual vectors* $\hat{r}_j$ *obtained by application of the GMRES method to the auxiliary system with* $\hat{\mathbf{A}} = \mathbf{A} - by^T$ *and initial guess* $\hat{x}_0 = 0$ *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 distance form b to the subspaces

$$\hat{\mathbf{A}}\mathcal{K}_j(\hat{\mathbf{A}}, b) = (\mathbf{A} - by^T)\mathcal{K}_j(\mathbf{A}, b),\ j = 1, \ldots, k,$$

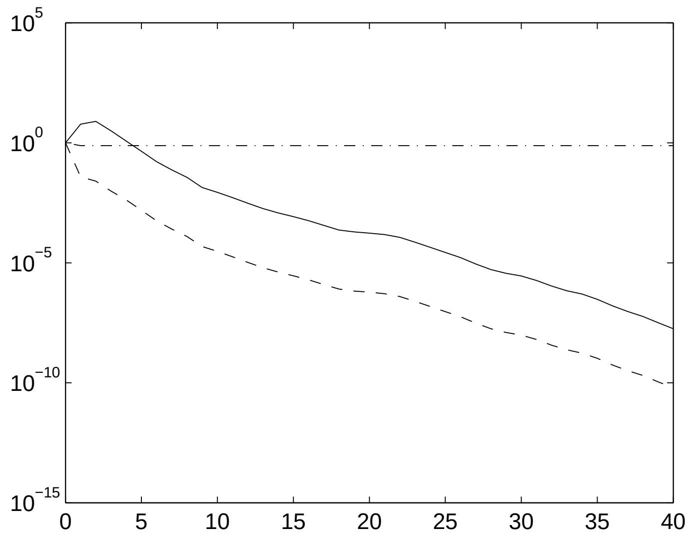
[Greenbaum, Strakoš - 94], which we can modify with the parameter vector y. Hence we can

1. construct a vector $y \in \mathbb{R}^n$ that prescribes the first k residual norms generated by GMRES applied to $(\mathbf{A} - by^T)\widehat{x} = b$ with initial guess $\widehat{x}_0 = 0$;

2. apply GMRES(m) to $(\mathbf{A} - by^T)\widehat{x} = b$ and presume that the first k residual norms of every restart decrease similarly as the prescribed norms.

Example: Linear system from a convection-diffusion problem (dimension 1225). We prescribe the first 10 residual norms of

GMRES(30): $\|r_1\| = 0.9$, $\|r_2\| = 0.8$, $\|r_3\| = 0.7$, $\|r_4\| = 0.6$, $\|r_5\| = 0.5$, $\|r_6\| = 0.4$, $\|r_7\| = 0.3$, $\|r_8\| = 0.2$, $\|r_9\| = 0.1$.



4. Open questions

- Normal case: The succeeding of our approach depends essentially on the quality of the used approximate eigenvalues. This holds in fact for many other GMRES accelerating techniques too.

- Non-normal case: More heuristical idea, less clear why the one stagnation can be overcome but the next can not. The influence of the choice of prescribed residual norms on the quality of the back-transformation has to be better understood. Surprisingly, the strategy works especially well for many linear systems arising from discretized partial differential equations, but we are not able to say why this is so.

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