Linear Algebra and its Applications 552 (2018) 219–238

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

On the convergence rate of DGMRES

Anne Greenbaum [a,*], Faranges Kyanfar [b], Abbas Salemi [b]

[a] *University of Washington, Applied Mathematics Department, Box 353925, Seattle, WA 98195, United States of America*
[b] *Shahid Bahonar University of Kerman, Applied Mathematics Department, Kerman, Iran*

ARTICLE INFO

Article history:
Received 1 September 2017
Accepted 22 April 2018
Available online 26 April 2018
Submitted by D.B. Szyld

MSC:
65F10
65F20
15A06

Keywords:
Drazin inverse
DGMRES
GMRES
Singular systems
Ritz values

ABSTRACT

The DGMRES algorithm was designed to compute the Drazin inverse solution of consistent or inconsistent linear systems $Ax = b$, where A is a square singular matrix with arbitrary index. In this paper we compare the convergence rate of DGMRES with that of GMRES applied to the nonsingular part of A.

* Corresponding author.
E-mail addresses: greenbau@uw.edu (A. Greenbaum), kyanfar@uk.ac.ir (F. Kyanfar), salemi@uk.ac.ir (A. Salemi).

https://doi.org/10.1016/j.laa.2018.04.027
0024-3795/

1. Introduction

Let A be an n by n matrix with index α. The index is the size of the largest Jordan block of A corresponding to the zero eigenvalue. The Drazin inverse A^D of A is the unique n by n matrix that satisfies

$$AA^D = A^D A, \quad A^{\alpha+1}A^D = A^\alpha, \quad A^D A A^D = A^D.$$

Since A^D can be written as a polynomial in A [8, p. 186], there is a possibility of using Krylov subspace methods to find the Drazin inverse solution $A^D b$ to a possibly inconsistent linear system $Ax = b$.

Such an algorithm, called DGMRES, was developed by Sidi in [13,14]. More recently, various restarted versions of the algorithm have been studied; see, for example, [11,18]. In this paper, we deal only with the full (unrestarted) version of the algorithm and compare its performance to that of the full GMRES algorithm applied to the nonsingular part of the matrix.

The DGMRES algorithm works as follows. Given an initial guess x_0, compute the initial residual $r_0 = b - Ax_0$, and form $A^\alpha r_0$. Note that while the original linear system $Ax = b$ may have no solution, if we multiply each side by A^α, then the linear system $A^{\alpha+1}x = A^\alpha b$ is consistent and has $x = A^D b$ as a solution.

Let $v_1 = \beta^{-1}(A^\alpha r_0)$, where $\beta = \|A^\alpha r_0\|$. (Here, and throughout this paper, $\|\cdot\|$ denotes the 2-norm for vectors and the corresponding operator norm for matrices.) We will let $x_\alpha = \cdots = x_1 = x_0$ and we will choose approximate solutions x_k, $k = \alpha+1, \alpha+2, \ldots$, to be of the form x_0 plus a linear combination of vectors from the Krylov subspace $\text{span}\{v_1, Av_1, \ldots, A^{k-\alpha-1}v_1\}$, where the linear combination is chosen to minimize $\|A^\alpha(b - Ax_k)\|$.

To do this, first use the Arnoldi algorithm to construct an orthonormal basis for the Krylov space $\text{span}\{v_1, Av_1, \ldots, A^{k-1}v_1\}$. If the Arnoldi vectors $v_1, \ldots, v_k$ form the columns of an n by k matrix V_k, then V_k satisfies

$$AV_k = V_{k+1}H_{k+1,k}, \tag{1}$$

where $H_{k+1,k}$ is a $k+1$ by k upper Hessenberg matrix. Take the approximate solution x_k to be of the form

$$x_k = x_0 + V_{k-\alpha}\xi,$$

for some $(k-\alpha)$-vector ξ to be determined. Then the residual $r_k := b - Ax_k$ satisfies $r_k = r_0 - AV_{k-\alpha}\xi$ and

$$A^\alpha r_k = A^\alpha r_0 - A^{\alpha+1}V_{k-\alpha}\xi.$$

Using (1), this can be written as

$$\begin{aligned} A^\alpha r_k &= A^\alpha r_0 - A^\alpha V_{k-\alpha+1} H_{k-\alpha+1,k-\alpha}\xi \\ &= A^\alpha r_0 - A^{\alpha-1} V_{k-\alpha+2} H_{k-\alpha+2,k-\alpha+1} H_{k-\alpha+1,k-\alpha}\xi \\ &\vdots \\ &= A^\alpha r_0 - A V_k H_{k,k-1} \cdots H_{k-\alpha+1,k-\alpha}\xi \\ &= A^\alpha r_0 - V_{k+1}\hat{H}_{k+1,k-\alpha}\xi, \end{aligned}$$

where $\hat{H}_{k+1,k-\alpha} = H_{k+1,k}H_{k,k-1}\cdots H_{k-\alpha+1,k-\alpha}$. Since $A^\alpha r_0 = V_{k+1}(\beta e_1)$, where $e_1 = (1,0,\ldots,0)^T$ is the first unit vector, we can write

$$A^\alpha r_k = V_{k+1}(\beta e_1 - \hat{H}_{k+1,k-\alpha}\xi), \tag{2}$$

and ξ can be chosen to minimize $\|A^\alpha r_k\|$ by solving the $k+1$ by $k-\alpha$ least squares problem

$$\hat{H}_{k+1,k-\alpha}\xi \approx \beta e_1.$$

Note that while $x_k - x_0$ is a linear combination of just the first $k-\alpha$ Arnoldi vectors, determination of the appropriate linear combination through the vector ξ requires computation of the first $k+1$ Arnoldi vectors and the associated Hessenberg matrices $H_{k+1,k}, H_{k,k-1}, \ldots$. Once a complete set of linearly independent Krylov space vectors has been constructed so that, say, $h_{K+1,K} = 0$ and $AV_K = V_K H_K$, the formula for $A^\alpha r_{K+\ell}$, $0 \le \ell \le \alpha$, becomes

$$\begin{aligned} A^\alpha r_{K+\ell} &= A^\alpha r_0 - A^{\alpha+1} V_{K+\ell-\alpha}\xi \\ &= A^\alpha r_0 - A^\alpha V_{K+\ell-\alpha+1} H_{K+\ell-\alpha+1,K+\ell-\alpha}\xi \\ &\vdots \\ &= A^\alpha r_0 - A^{\ell+1} V_K \hat{H}_{K,K+\ell-\alpha}\xi \\ &= A^\alpha r_0 - V_K H_K^{\ell+1}\hat{H}_{K,K+\ell-\alpha}\xi \\ &= V_K(\beta e_1 - H_K^{\ell+1}\hat{H}_{K,K+\ell-\alpha}\xi), \end{aligned}$$

and a K by $K+\ell-\alpha$ least squares problem with coefficient matrix $H_K^{\ell+1}\hat{H}_{K,K+\ell-\alpha}$ must be solved to determine ξ. When $\ell = \alpha$, the least squares problem involves the square matrix $H_K^{\alpha+1}$ so that if this matrix is nonsingular then an exact solution to $A^{\alpha+1}x = A^\alpha b$ is obtained.

This algorithm is very similar to the GMRES algorithm for solving nonsingular linear systems, and it can be written as in Table 1. Note from (2) that one can monitor $\|A^\alpha r_k\| = \|\hat{H}_{k+1,k-\alpha}\xi - \beta e_1\|$, without actually computing x_k, so one need only evaluate x_k when this norm drops below a desired tolerance.

Table 1
DGMRES Algorithm.

DGMRES for the Drazin inverse solution of $Ax = b$, index of A is α. [14]

Initialization:
Given an initial guess x_0, compute $r_0 = b - Ax_0$ and $A^\alpha r_0$. Set $\beta = \|A^\alpha r_0\|$ and $v_1 = \beta^{-1} A^\alpha r_0$.
Let $x_1 = \cdots = x_\alpha = x_0$.
Main loop:
For $k = 1, 2, \ldots$

Construct the next Arnoldi vector: Set $v_{k+1} = Av_k$. For $j = 1, \ldots, k$:
Compute $h_{jk} = v_j^* v_{k+1}$.
Replace $v_{k+1} \leftarrow v_{k+1} - h_{jk} v_j$.
Compute $h_{k+1,k} = \|v_{k+1}\|$.
If $h_{k+1,k} = 0$ (or within roundoff of 0), set a flag; otherwise set $v_{k+1} = v_{k+1}/h_{k+1,k}$.

If $k > \alpha$ and flag not set,
Form $\hat{H}_{k+1,k-\alpha} = \prod_{j=0}^{\alpha} H_{k+1-j,k-j}$.
Solve the least squares problem $\hat{H}_{k+1,k-\alpha}\xi \approx \beta e_1$ for ξ.
Set $x_k = x_0 + V_{k-\alpha}\xi$. Compute $r_k = b - Ax_k$ and $A^\alpha r_k$.
If $\|A^\alpha r_k\|$ sufficiently small, then terminate; otherwise, continue.

If flag is set, form the final iterates and terminate: For $\ell = \max\{0, \alpha + 1 - k\}, \ldots, \alpha$,
Form $\hat{H}_{k,k+\ell-\alpha} = \prod_{j=0}^{\alpha-\ell-1} H_{k-j,k-j-1}$ and $\hat{\hat{H}} = H_k^{\ell+1} \hat{H}_{k,k+\ell-\alpha}$.
Solve the least squares problem $\hat{\hat{H}}\xi \approx \beta e_1$ for ξ.
Set $x_{k+\ell} = x_0 + V_{k+\ell-\alpha}\xi$. Compute $r_{k+\ell} = b - Ax_{k+\ell}$ and $A^\alpha r_{k+\ell}$.

Note also that if α in the algorithm of Table 1 is greater than the index of A, then the procedure still finds a solution of $A^{\alpha+1}x = A^\alpha b$, while if α is less than the index of A, then it attempts to solve the linear system $A^{\alpha+1}x = A^\alpha b$, which may have no solution.

In the following sections, we compare the convergence rate of the DGMRES algorithm to that of GMRES applied to the nonsingular part of the matrix. It is shown that while the GMRES algorithm produces a residual at step k that is equal to the initial residual $\hat{r}_0$ minus its orthogonal projection onto $\text{span}\{B\hat{r}_0, B^2\hat{r}_0, \ldots, B^k\hat{r}_0\}$, where B is the nonsingular part of the matrix, the convergence of DGMRES for $k > \alpha$ is governed by the size of $\hat{r}_0$ minus its orthogonal projection onto $\text{span}\{B^{\alpha+1}\hat{r}_0, B^{\alpha+2}\hat{r}_0, \ldots, B^k\hat{r}_0\}$. Thus, the problem is to estimate the importance of the missing terms $B\hat{r}_0, \ldots, B^\alpha\hat{r}_0$. Good *a priori* estimates can be given when $\alpha = 1$. The same techniques can be used for larger values of α, but the bounds become more complicated and less tight.

2. Expression for the Drazin inverse using a block upper triangular form

Every n by n matrix A can be written in the form

$$A = S \begin{bmatrix} \mathscr{B} & 0 \\ 0 & \mathscr{N} \end{bmatrix} S^{-1}, \tag{3}$$

where S is a nonsingular n by n matrix, $\mathscr{B}$ is a nonsingular m by m matrix, $0 \le m \le n$, and $\mathscr{N}$ is a nilpotent $n - m$ by $n - m$ matrix [8, p. 185]. This could, for example, be the Jordan canonical form of A, where $\mathscr{B}$ is a direct sum of all the nonsingular Jordan

blocks and $\mathscr{N}$ is a direct sum of all the singular Jordan blocks. The Drazin inverse of A is then given by

$$A^D = S \begin{bmatrix} \mathscr{B}^{-1} & 0 \\ 0 & 0 \end{bmatrix} S^{-1}. \tag{4}$$

Note that while the decomposition (3) is not unique, the Drazin inverse (4) is.

We prefer to work with unitary similarity transformations, so factoring the nonsingular matrix S in the form $S = QR$ where Q is unitary and R is upper triangular of the form

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix},$$

where R_{11} is a nonsingular m by m matrix and R_{22} is a nonsingular $n-m$ by $n-m$ matrix, we can write

$$A = Q \begin{bmatrix} B & * \\ 0 & N \end{bmatrix} Q^*, \tag{5}$$

where

$$B = R_{11} \mathscr{B} R_{11}^{-1}, \quad N = R_{22} \mathscr{N} R_{22}^{-1},$$

and the asterisk represents a matrix block that will not play a role in the analysis. Note that if the index of A is α, then the first m columns of Q form an orthonormal basis for the range of A^α and the remaining $n-m$ columns of Q form an orthonormal basis for the orthogonal complement of the range of A^α.

While the decompositions (3) or (5) are not easy to compute numerically (see, for example, [7] for a method to compute (5)), in the next section we will use the form (5) to compare the performance of DGMRES applied to A to that of GMRES applied to B.

3. Comparison with GMRES

Assume that A has index α. The DGMRES algorithm described in Section 1 generates vectors x_k^D, $k = \alpha+1, \alpha+2, \ldots$, of the form

$$x_k^D = x_0 + \sum_{j=1}^{k-\alpha} c_j A^{\alpha+j-1} r_0.$$

(Here we have added a superscript D to emphasize that these are quantities generated by the DGMRES algorithm.) The residual $r_k^D := b - Ax_k^D$ is then $r_0 - \sum_{j=1}^{k-\alpha} c_j A^{\alpha+j} r_0$, and $A^\alpha r_k^D$ satisfies

$$A^\alpha r_k^D = (I - \sum_{j=1}^{k-\alpha} c_j A^{\alpha+j}) A^\alpha r_0.$$

The coefficients $c_1, \ldots, c_{k-\alpha}$ are chosen to minimize the 2-norm of $A^\alpha r_k^D$:

$$\|A^\alpha r_k^D\| = \min_{c_1,\ldots,c_{k-\alpha}} \|(I - \sum_{j=1}^{k-\alpha} c_j A^{\alpha+j}) A^\alpha r_0\|. \tag{6}$$

From (5) and (6),

$$\|A^\alpha r_k^D\| = \min_{c_1,\ldots,c_{k-\alpha}} \left\| Q \begin{bmatrix} I - \sum_{j=1}^{k-\alpha} c_j B^{\alpha+j} & * \\ 0 & I \end{bmatrix} Q^* A^\alpha r_0 \right\|. \tag{7}$$

The vector $Q^* A^\alpha r_0$ has norm equal to $\|A^\alpha r_0\|$ and it has nonzeros only in its first m positions since columns $m+1, \ldots, n$ of Q are orthogonal to the range of A^α. If $\hat{r}_0$ denotes the first m entries of this vector, then

$$\|A^\alpha r_k^D\| = \min_{c_1,\ldots,c_{k-\alpha}} \|(I - \sum_{j=1}^{k-\alpha} c_j B^{\alpha+j}) \hat{r}_0\|, \tag{8}$$

and, for any initial residual r_0, we can write

$$\|A^\alpha r_k^D\| / \|A^\alpha r_0\| \le \max_{\|w\|=1} \min_{c_1,\ldots,c_{k-\alpha}} \|(I - \sum_{j=1}^{k-\alpha} c_j B^{\alpha+j}) w\| \tag{9}$$

$$\le \min_{c_1,\ldots,c_{k-\alpha}} \|I - \sum_{j=1}^{k-\alpha} c_j B^{\alpha+j}\|. \tag{10}$$

The bound in (9) is the worst-case residual norm bound for DGMRES while the one in (10) is the *ideal DGMRES* residual norm bound [6]. It is clear that the right-hand side of (9) is less than or equal to that in (10) since for any vector w with $\|w\| = 1$, we have $\min_{c_1,\ldots,c_{k-\alpha}} \|(I - \sum_{j=1}^{k-\alpha} c_j B^{\alpha+j}) w\| \le \min_{c_1,\ldots,c_{k-\alpha}} \|I - \sum_{j=1}^{k-\alpha} c_j B^{\alpha+j}\|$, but examples are given in [4,16] of cases where these two quantities are *not* equal. It follows from results in [9] that the coefficients $c_1, \ldots, c_{k-\alpha}$ that minimize the matrix norm in (10) are uniquely determined, assuming that this minimal matrix norm is positive.

When the GMRES algorithm is applied to a problem with the nonsingular coefficient matrix B and initial residual $\hat{r}_0$, the residual r_k^G at step k satisfies

$$\|r_k^G\| = \min_{d_1,\ldots,d_k} \|(I - \sum_{j=1}^{k} d_j B^j) \hat{r}_0\|. \tag{11}$$

The worst-case and *ideal* GMRES residual norm bounds are:

$$\|r_k^G\|/\|r_0^G\| \le \max_{\|w\|=1} \min_{d_1,\ldots,d_k} \|(I - \sum_{j=1}^{k} d_j B^j)w\| \tag{12}$$

$$\le \min_{d_1,\ldots,d_k} \|I - \sum_{j=1}^{k} d_j B^j\|. \tag{13}$$

The terms $d_1 B, \ldots, d_\alpha B^\alpha$ that are present in the GMRES minimization problems (11)–(13) are *not* present in the DGMRES minimization problems (8)–(10), and we wish to estimate how important these extra terms are.

We first note that it is immediate from (8) that

$$\|A^\alpha r_k^D\| \le \min_{d_1,\ldots d_{\lfloor k/(\alpha+1)\rfloor}} \|(I - \sum_{j=1}^{\lfloor (k/(\alpha+1)\rfloor} d_j (B^{\alpha+1})^j)\hat{r}_0\|, \tag{14}$$

where $\lfloor\cdot\rfloor$ denotes the integer part. The quantity on the right is the norm of the GMRES residual at step $\lfloor k/(\alpha+1)\rfloor$ for a problem with coefficient matrix $B^{\alpha+1}$ and initial residual $\hat{r}_0$. We wish, however, to relate the behavior of DGMRES for a problem with coefficient matrix A to that of GMRES for a problem with coefficient matrix B rather than $B^{\alpha+1}$.

3.1. Matrices of index one

Consider first the case $\alpha = 1$. Let $\tilde{d}_1, \ldots, \tilde{d}_k$ be the coefficients that achieve the minimum in (11). Then we can write

$$\hat{r}_0 = \sum_{j=1}^{k} \tilde{d}_j B^j \hat{r}_0 + r_k^G, \quad B\hat{r}_0 = \sum_{j=1}^{k} \tilde{d}_j B^{j+1} \hat{r}_0 + B r_k^G.$$

Combining these two equalities,

$$\hat{r}_0 = \tilde{d}_1 \left(\sum_{j=1}^{k} \tilde{d}_j B^{j+1} \hat{r}_0 + B r_k^G \right) + \sum_{j=1}^{k-1} \tilde{d}_{j+1} B^{j+1} \hat{r}_0 + r_k^G$$

$$= \sum_{j=1}^{k} (\tilde{d}_1 \tilde{d}_j + \tilde{d}_{j+1}) B^{j+1} \hat{r}_0 + \tilde{d}_1 B r_k^G + r_k^G, \quad \tilde{d}_{k+1} := 0. \tag{15}$$

Thus

$$\hat{r}_0 - \sum_{j=1}^{k} (\tilde{d}_1 \tilde{d}_j + \tilde{d}_{j+1}) B^{j+1} \hat{r}_0 = (I + \tilde{d}_1 B) r_k^G,$$

and it follows from (8) that

$$\|Ar_{k+1}^D\| = \min_{c_1,\ldots,c_k} \|\hat{r}_0 - \sum_{j=1}^{k} c_j B^{1+j}\hat{r}_0\| \le \|(I + \tilde{d}_1 B)r_k^G\| \le (1 + |\tilde{d}_1|\ \|B\|)\|r_k^G\|. \quad (16)$$

A slightly stronger bound can be obtained as follows. Comparing (8) and (11) with k replaced by $k+1$, we can write

$$\begin{aligned}
&\min_{c_1,\ldots,c_k} \|(I - \sum_{j=1}^{k} c_j B^{1+j})\hat{r}_0\| \\
&= \min_{d_1,\ldots,d_{k+1},e_1,\ldots,e_k} \|(I - \sum_{j=1}^{k+1} d_j B^j)\hat{r}_0 + (d_1 B - \sum_{j=1}^{k} e_j B^{1+j})\hat{r}_0\| \\
&\le \min_{d_1,\ldots,d_{k+1}} \|(I - \sum_{j=1}^{k+1} d_j B^j)\hat{r}_0\| + \min_{e_1,\ldots,e_k} \|(\tilde{d}_1 B - \sum_{j=1}^{k} e_j B^{1+j})\hat{r}_0\| \\
&= \min_{d_1,\ldots,d_{k+1}} \|(I - \sum_{j=1}^{k+1} d_j B^j)\hat{r}_0\| + \\
&\quad |\tilde{d}_1| \min_{e_1,\ldots,e_k} \|B(I - \sum_{j=1}^{k} e_j B^j)\hat{r}_0\|, \qquad (17)
\end{aligned}$$

where $\tilde{d}_1$ is the value of d_1 that minimizes the first term on the right-hand side. The first term on the right-hand side is the norm of the GMRES residual r_{k+1}^G, and the second measures how well $\tilde{d}_1 B\hat{r}_0$ can be approximated by a linear combination of $B^2\hat{r}_0, \ldots, B^{k+1}\hat{r}_0$. It is $|\tilde{d}_1|$ times the norm of the GMRES residual at step k when the initial residual is $B\hat{r}_0$, and it is bounded above by $|\tilde{d}_1|\ \|B\|$ times the norm of the GMRES residual at step k for initial residual $\hat{r}_0$. Thus, letting r_{k+1}^D denote the DGMRES residual at step $k+1$, we can write

$$\|Ar_{k+1}^D\| \le \|r_{k+1}^G\| + |\tilde{d}_1|\ \|B\|\ \|r_k^G\|. \quad (18)$$

Similar comparisons can be derived for worst-case and ideal GMRES and DGMRES. For instance, if $P_{k+1}^D(B)$ represents the ideal DGMRES polynomial in (10) (of degree $k+1$ but involving only the k powers $B^2, \ldots, B^{k+1}$) and $P_k^G(B)$ represents the ideal GMRES polynomial in (13), then

$$\|P_{k+1}^D(B)\| \le \|P_{k+1}^G(B)\| + |\tilde{\tilde{d}}_1|\ \|B\|\ \|P_k^G(B)\|, \quad (19)$$

where $\tilde{\tilde{d}}_1$ is the coefficient of B in $P_{k+1}^G(B)$.

The only remaining question is the size of $\tilde{d}_1$ or $\tilde{\tilde{d}}_1$, which are the coefficients of B in the actual and ideal GMRES polynomials, respectively. It is well-known that the reciprocals of the roots of the GMRES polynomial corresponding to any initial residual $\hat{r}_0$ all lie in

the field of values of B^{-1}: $W(B^{-1}) := \{q^*B^{-1}q : \|q\| = 1\}$. [See, e.g., [15,5].] It is shown in [17] that this holds for the reciprocals of the roots of the ideal GMRES polynomial as well. If $w(B^{-1})$ denotes the numerical radius of B^{-1}, $w(B^{-1}) := \max_{z\in W(B^{-1})} |z|$, then the reciprocal of each root has absolute value less than or equal to $w(B^{-1})$. If $\theta_1, \ldots, \theta_{k+1}$ are the roots of the actual or ideal GMRES polynomial, then this polynomial can be written as $\prod_{j=1}^{k+1}(I - \theta_j^{-1}B)$, and the coefficient of B is the negative of the sum of the reciprocals of the roots, which satisfies $|-\sum_{j=1}^{k+1}\theta_j^{-1}| \le (k+1)\, w(B^{-1})$. It follows that the bounds (18) and (19), which depend on coefficients generated by the GMRES algorithm or its ideal counterpart, can be replaced by the *a priori* estimates

$$\|Ar_{k+1}^D\| \le \|r_{k+1}^G\| + (k+1)\; w(B^{-1})\; \|B\|\; \|r_k^G\| \le \|r_{k+1}^G\| + (k+1)\; \kappa(B)\; \|r_k^G\|, \quad (20)$$

and

$$\begin{aligned} \|P_{k+1}^D(B)\| &\le \|P_{k+1}^G(B)\| + (k+1)\; w(B^{-1})\; \|B\|\; \|P_k^G(B)\| \\ &\le \|P_{k+1}^G(B)\| + (k+1)\; \kappa(B)\; \|P_k^G(B)\|, \end{aligned} \quad (21)$$

respectively, where $\kappa(B) := \|B^{-1}\|\; \|B\|$. The second bound in (20) and (21) is just a slightly weaker (but possibly more convenient) estimate using the fact that $w(B^{-1}) \le \|B^{-1}\|$.

Unfortunately, the bounds in (20) and (21) involve a factor $k+1$ in the estimates of $|\tilde{d}_1|$ and $|\tilde{\tilde{d}}_1|$. We need a bound on the absolute value of the sum of the reciprocals of the roots of the actual or ideal GMRES polynomial. Could all $k+1$ of these values actually be equal to $w(B^{-1})$? It seems unlikely, but in the case of the ideal GMRES polynomial we know of no further published results on the distribution of the reciprocal roots.

In the case of actual GMRES polynomials, however, there is some analysis [3,10]. The coefficient estimates in [10] depend on the initial residual, but to analyze worst-case GMRES we need bounds that are independent of the starting vector. Such *a priori* bounds can be found in [3], which deals with sums of real parts and with absolute values of Ritz values, but the same arguments can be used to obtain bounds on sums of imaginary parts and hence on the absolute value of the sum of Ritz values.

The roots of the GMRES polynomial p_{k+1}^G corresponding to the nonsingular matrix B and some initial residual $\hat{r}_0$ are known as *harmonic Ritz values*. It is shown [5] that the reciprocals of these roots are standard Ritz values obtained from the Arnoldi process applied to the matrix B^{-1} with starting vector $B\hat{r}_0$. Write B^{-1} in the form $B^{-1} = H + iS$, where $H := (B^{-1} + B^{-*})/2$ and $S := (B^{-1} - B^{-*})/(2i)$ are both Hermitian. Let $\mu_1 \le \ldots \le \mu_m$ denote the eigenvalues of H and let $\nu_1 \le \ldots \le \nu_m$ denote the eigenvalues of S. Let $\theta_1^{-1}, \ldots, \theta_{k+1}^{-1}$ denote the Ritz values obtained at step $k+1$ of the Arnoldi algorithm applied to B^{-1} with initial vector $B\hat{r}_0$ (i.e., the reciprocals of the roots of the GMRES polynomial p_{k+1}^G).

Theorem 1. *(See [3].) Using the above notation, if $\tilde{d}_1$ is the coefficient of B in the GMRES polynomial $p_{k+1}^G(B)$ corresponding to the nonsingular matrix B and initial residual $\hat{r}_0$ (i.e., the polynomial for which $r_{k+1}^G = p_{k+1}^G(B)\hat{r}_0$), then*

$$|\tilde{d}_1| = \left| \sum_{j=1}^{k+1} \theta_j^{-1} \right| \leq \tag{22}$$

$$\left[\left(\max\{ |\sum_{j=1}^{k+1} \mu_{m-j+1}|, \ |\sum_{j=1}^{k+1} \mu_j| \} \right)^2 + \left(\max\{ |\sum_{j=1}^{k+1} \nu_{m-j+1}|, \ |\sum_{j=1}^{k+1} \nu_j| \} \right)^2 \right]^{1/2}.$$

Proof. Let the columns of an m by $k+1$ matrix V form an orthonormal basis for the Krylov space generated by the Arnoldi algorithm applied to B^{-1} with initial vector $B\hat{r}_0$. Then V can be chosen in such a way that $V^*B^{-1}V$ is upper triangular, with the Ritz values $\theta_1^{-1}, \ldots, \theta_{k+1}^{-1}$ on its main diagonal. Let the columns of $\hat{V} \in \mathbb{C}^{m\times(m-k-1)}$ form an orthonormal basis for the orthogonal complement of the range of V, and choose this basis so that $\hat{V}^*B^{-1}\hat{V}$ is upper triangular, with diagonal entries labeled $\theta_{k+2}^{-1}, \ldots, \theta_m^{-1}$. Order the values $\theta_1^{-1}, \ldots, \theta_m^{-1}$ in increasing order of real part: $\mathrm{Re}(\theta_{r_1}^{-1}) \leq \ldots \leq \mathrm{Re}(\theta_{r_m}^{-1})$, or in increasing order of imaginary part: $\mathrm{Im}(\theta_{i_1}^{-1}) \leq \ldots \leq \mathrm{Im}(\theta_{i_m}^{-1})$.

By a result of Schur [1, p. 35], the vector $[\mathrm{Re}(\theta_{r_j}^{-1})]_{j=1}^m$ of ordered diagonal entries of $[V, \hat{V}]^*H[V, \hat{V}]$ majorizes the vector $[\mu_j]_{j=1}^m$ of eigenvalues; that is,

$$\sum_{j=1}^{\ell} \mu_j \leq \sum_{j=1}^{\ell} \mathrm{Re}(\theta_{r_j}^{-1}), \quad \ell = 1, \ldots, m. \tag{23}$$

Since the values $\theta_{r_j}^{-1}$, $j = 1, \ldots, k+1$ are ordered by increasing real part, it follows that $\sum_{j=1}^{\ell} \mathrm{Re}(\theta_{r_j}^{-1}) \leq \sum_{j=1}^{\ell} \mathrm{Re}(\theta_j^{-1})$, and (23) can be replaced by

$$\sum_{j=1}^{k+1} \mu_j \leq \sum_{j=1}^{k+1} \mathrm{Re}(\theta_j^{-1}). \tag{24}$$

Similarly, the vector $[\mathrm{Im}(\theta_{i_j}^{-1})]_{j=1}^m$ of ordered diagonal entries of $[V, \hat{V}]^*S[V, \hat{V}]$ majorizes the vector $[\nu_j]_{j=1}^m$ of eigenvalues; that is,

$$\sum_{j=1}^{\ell} \nu_j \leq \sum_{j=1}^{\ell} \mathrm{Im}(\theta_{i_j}^{-1}), \quad \ell = 1, \ldots, m. \tag{25}$$

Since the values $\theta_{i_j}^{-1}$, $j = 1, \ldots, k+1$ are ordered by increasing imaginary part, it follows that $\sum_{j=1}^{\ell} \mathrm{Im}(\theta_{i_j}^{-1}) \leq \sum_{j=1}^{\ell} \mathrm{Im}(\theta_j^{-1})$ and so we have

$$\sum_{j=1}^{k+1} \nu_j \le \sum_{j=1}^{k+1} \mathrm{Im}(\theta_j^{-1}). \tag{26}$$

Applying the same analysis to $-B^{-1}$ yields

$$\sum_{j=1}^{k+1} \mu_{m-j+1} \ge \sum_{j=1}^{k+1} \mathrm{Re}(\theta_j^{-1}), \quad \sum_{j=1}^{k+1} \nu_{m-j+1} \ge \sum_{j=1}^{k+1} \mathrm{Im}(\theta_j^{-1}). \tag{27}$$

Finally, combining (24), (26), and (27), we obtain the bounds,

$$\left| \sum_{j=1}^{k+1} \mathrm{Re}(\theta_j^{-1}) \right| \le \max \left\{ |\sum_{j=1}^{k+1} \mu_{m-j+1}|, \ |\sum_{j=1}^{k+1} \mu_j| \right\},$$
$$\left| \sum_{j=1}^{k+1} \mathrm{Im}(\theta_j^{-1}) \right| \le \max \left\{ |\sum_{j=1}^{k+1} \nu_{m-j+1}|, \ |\sum_{j=1}^{k+1} \nu_j| \right\},$$

from which (22) follows. □

If many eigenvalues of H and S have magnitude significantly less than $w(B^{-1})$ or if their signs cause cancellation in the sums on the right-hand side of (22), then the bound (22) may be significantly less than $(k+1)w(B^{-1})$.

For worst-case DGMRES, if $\|Ar_{k+1}^D\|$ is the maximum, over all initial residuals r_0 with $\|Ar_0\| = 1$, of the norm of A times the residual at step $k+1$, and if $\|r_{k+1}^G\|$ and $\|r_k^G\|$ are the GMRES residual norms at steps $k+1$ and k for the associated problem with coefficient matrix B and initial residual $\hat{r}_0$ consisting of the first m entries of $Q^* Ar_0$, then, in analogy to (18), we can write

$$\max_{\|Ar_0\|=1} \|Ar_{k+1}^D\| \le \|r_{k+1}^G\| + |\tilde{d}_1| \ \|B\| \ \|r_k^G\|, \tag{28}$$

where now $\tilde{d}_1$ is the coefficient of B in the GMRES polynomial $p_{k+1}^G(B)$ for initial residual $\hat{r}_0$ corresponding to the worst-case DGMRES initial residual. The bound $|\tilde{d}_1| \le (k+1)w(B^{-1})$ holds in this case as well, but we also have the sharper estimate in (22).

Note also that if A has index 0 (i.e., it is nonsingular), but the DGMRES algorithm is run with $\alpha = 1$, then these same comparisons hold. In this case $B = A$, but because we treat the matrix as if it had index 1 in the DGMRES algorithm the first degree term is still missing from the DGMRES polynomial. The following theorem summarizes these results:

Theorem 2. *Let A be an n by n matrix with index 0 or 1 and assume that A can be written in the form (5), where B is a nonsingular m by m matrix, $m \le n$. For $k \ge 1$, let r_{k+1}^D denote the residual at step $k+1$ of the DGMRES algorithm applied to the linear*

system $Ax = b$ with initial residual $r_0 := b - Ax_0$ and with α set to 1. Let r^G_{k+1} and r^G_k denote the residuals at steps $k+1$ and k of the GMRES algorithm applied to a linear system with coefficient matrix B and initial residual $\hat{r}_0$ consisting of the first m entries of Q^*Ar_0. Then

$$\|Ar^D_{k+1}\| \le \|r^G_{k+1}\| + |\tilde{d}_1| \, \|B\| \, \|r^G_k\|, \tag{29}$$

where $\tilde{d}_1$ is the coefficient of B in the GMRES polynomial $p^G_{k+1}(B)$, where $r^G_{k+1} = p^G_{k+1}(B)\hat{r}_0$. The coefficient $\tilde{d}_1$ satisfies $|\tilde{d}_1| \le (k+1)w(B^{-1})$, where $w(\cdot)$ denotes the numerical radius, and it also satisfies the inequality in (22).

If $P^D_{k+1}(B)$ is the ideal DGMRES polynomial defined by (10) and P^G_{k+1} and P^G_k are ideal GMRES polynomials defined by (13), then

$$\|P^D_{k+1}(B)\| \le \|P^G_{k+1}(B)\| + |\tilde{\tilde{d}}_1| \, \|B\| \, \|P^G_k(B)\|, \tag{30}$$

where $\tilde{\tilde{d}}_1$ is the coefficient of B in the ideal GMRES polynomial $P^G_{k+1}(B)$ and satisfies $|\tilde{\tilde{d}}_1| \le (k+1)w(B^{-1})$.

3.2. Matrices with higher index

While the same procedure can be applied to matrices of higher index, its usefulness becomes questionable. Suppose, for instance, that $\alpha = 2$. Returning, for simplicity, to the type of argument that led to (16), let $\mathscr{P}_2$ denote any element in the span of $\{B^3\hat{r}_0, \ldots, B^{k+2}\hat{r}_0\}$. Then we can write

$$\hat{r}_0 = \tilde{d}_1 B\hat{r}_0 + \tilde{d}_2 B^2\hat{r}_0 + r^G_k + \mathscr{P}_2, \quad B\hat{r}_0 = \tilde{d}_1 B^2\hat{r}_0 + Br^G_k + \mathscr{P}_2, \quad B^2\hat{r}_0 = B^2 r^G_k + \mathscr{P}_2,$$

where $\tilde{d}_1$ and $\tilde{d}_2$ are the coefficients of $B\hat{r}_0$ and $B^2\hat{r}_0$, respectively, in the GMRES polynomial p^G_k satisfying $r^G_k = p^G_k(B)\hat{r}_0$. Combining these equalities,

$$\begin{aligned} \hat{r}_0 &= \tilde{d}_1(\tilde{d}_1 B^2\hat{r}_0 + Br^G_k) + \tilde{d}_2 B^2 r^G_k + r^G_k + \mathscr{P}_2 \\ &= [I + \tilde{d}_1 B + (\tilde{d}_1^2 + \tilde{d}_2)B^2]r^G_k + \mathscr{P}_2. \end{aligned} \tag{31}$$

Since $\|A^2 r^D_{k+2}\|$ is the minimum over all elements $\mathscr{P}_2$ in $\text{span}\{B^3\hat{r}_0, \ldots, B^{k+2}\hat{r}_0\}$ of the norm of the difference between $\hat{r}_0$ and $\mathscr{P}_2$, it follows that

$$\|A^2 r^D_{k+2}\| \le \|(I + \tilde{d}_1 B + (\tilde{d}_1^2 + \tilde{d}_2)B^2)r^G_k\| \le \left(1 + |\tilde{d}_1| \, \|B\| + |\tilde{d}_1^2 + \tilde{d}_2| \, \|B^2\|\right) \|r^G_k\|. \tag{32}$$

Writing the GMRES polynomial in the form $\prod_{j=1}^k (I - \theta_j^{-1}B)$, it can be seen that the coefficient $\tilde{d}_2$ of B^2 is the sum of products of all pairs of reciprocal roots: $\tilde{d}_2 = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} \theta_i^{-1}\theta_j^{-1}$. There are $\frac{k(k-1)}{2}$ such pairs, each of which could be as large as $w(B^{-1})^2$ (although the results of Theorem 3.1 might preclude all of the reciprocal roots

from being equal to $w(B^{-1})$), so it is difficult to rule out the possibility that the term $|\tilde{d}_1^2 + \tilde{d}_2| \, \|B^2\|$ in (32) could be quite large.

One thing that *can* be said about DGMRES convergence is that if GMRES applied to a problem with coefficient matrix B and initial residual $\hat{r}_0$ obtains the exact solution after K steps, then DGMRES applied to the corresponding index α linear system $Ax = b$ obtains an exact solution to $A^{\alpha+1}x = A^\alpha b$ after $K + \alpha$ steps. [This is the Drazin inverse solution $A^D b$ if the projection of x_0 onto the null space of A^α is 0, which is certainly the case if $x_0 = 0$. See [13,14].]

Theorem 3. *Let A be an n by n matrix with index less than or equal to α and assume that A can be written in the form (5), where B is a nonsingular m by m matrix, $m \leq n$. Suppose the GMRES algorithm applied to a linear system with coefficient matrix B and initial residual $\hat{r}_0$ consisting of the first m entries of $Q^* A^\alpha r_0$ obtains the exact solution after K steps. Then DGMRES applied to the linear system $Ax = b$ obtains an exact solution to $A^{\alpha+1}x = A^\alpha b$ after $K + \alpha$ steps.*

Proof. Since GMRES obtains the exact solution after K steps, it follows from (11) that

$$\hat{r}_0 \in \text{span}\{B\hat{r}_0, B^2\hat{r}_0, \ldots, B^K\hat{r}_0\}. \tag{33}$$

To show that DGMRES finds an exact solution to $A^{\alpha+1}x = A^\alpha b$ after K steps, we must show, based on (8), that $\hat{r}_0$ lies in

$$\text{span}\{B^{\alpha+1}\hat{r}_0, \ldots, B^{\alpha+K}\hat{r}_0\}. \tag{34}$$

For $\alpha = 0$, this is immediate. Otherwise, multiplying (33) by B^α, then $B^{\alpha-1}$, etc., we find that $B^\alpha \hat{r}_0$ is in the subspace defined in (34), as is $B^{\alpha-1}\hat{r}_0$, ..., as is $B\hat{r}_0$. Therefore $\hat{r}_0$ lies in this subspace. □

Note that the converse of Theorem 3 is *not* true. If DGMRES obtains the exact solution after $K + \alpha$ steps (that is, if $\hat{r}_0$ lies in the space (34)), it does *not* follow that GMRES finds the exact solution after K steps (i.e., that (33) holds). A simple example is the companion matrix

$$B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix},$$

whose minimal polynomial is $z^3 - z^2 - 1$. Since $I = -B^2 + B^3$, it follows that DGMRES with $\alpha = 1$ obtains the exact Drazin inverse solution to any linear system with coefficient matrix equal to the direct sum of B and a nilpotent matrix of index 1 in two steps. Yet, if $\hat{r}_0 = e_1$, for instance, then the GMRES algorithm makes no progress in the first two steps. Of course, if DGMRES finds the exact solution after $K + \alpha$ steps, then GMRES does as well since it minimizes over a space that contains (34).

4. Numerical results

4.1. Neumann problem for Poisson's equation

Fig. 1 shows the convergence of DGMRES applied to a 5-point difference approximation to the problem

$$-\Delta u = f \text{ in } [0,1] \times [0,1],$$

with homogeneous Neumann boundary conditions: $u_x = 0$ when $x = 0$ or $x = 1$, $u_y = 0$ when $y = 0$ or $y = 1$. The finite difference matrix has zero row sums; that is, the vector of all 1's lies in its null space, corresponding to the fact that the solution is determined only up to an additive constant. A grid with $h = 1/100$ was used, resulting in a 99^2 by 99^2 matrix equation for the solution at the interior nodes, using the approximation that solution values at boundary nodes are equal to those one line in. A random right-hand side f was used, with a zero initial guess. The DGMRES convergence curve is marked with ∘'s, and the convergence curve for GMRES applied to the nonsingular part (B in (5)) is plotted with a solid line. Also plotted, with a dash–dot line, is the upper bound (18) using the computed value of $|\tilde{d}_1|$. The dashed line in the figure is the rightmost upper bound in (20), and the dotted line is the upper bound obtained from (18) when $|\tilde{d}_1|$ is replaced by the bound in (22). Note that the upper bound in (22) is crude initially, but as k increases it approximates $|\tilde{d}_1|$ well. The condition number of B is 7943.1, and it can be seen from the figure that the factor $(k+1)\,\kappa(B)$ in (20) is a significant overestimate of $|\tilde{d}_1|\,\|B\|$ in (18). Still, all curves show the same general convergence pattern.

Like GMRES, the DGMRES algorithm can be used with a nonsingular preconditioner to improve convergence. Preconditioning can be done on the left or the right or on both sides, resulting effectively in the DGMRES algorithm being applied to the linear system $L_1^{-1}AL_2^{-1}(L_2x) = L_1^{-1}b$. Based on the analysis of the previous section, one might expect faster convergence from DGMRES applied to the preconditioned system if the nonsingular part of the preconditioned matrix is better suited for fast GMRES convergence.

Although the matrix for the Neumann problem is singular, it has a nonsingular incomplete Cholesky decomposition. That is, there is a nonsingular lower triangular matrix L with the same sparsity pattern as the lower triangle of A such that LL^T matches A in places where A has nonzeros but has some nonzero entries in places where A has zeros. Applying DGMRES to the linear system $L^{-1}AL^{-T}(L^Tx) = L^{-1}b$, again with a random right-hand side b and a zero initial guess resulted in the convergence curve marked with ∘'s in Fig. 2. The solid curve shows the convergence of GMRES applied to the nonsingular part of $L^{-1}AL^{-T}$, and the dash–dot and dashed lines in the figure show the upper bounds (18) and (20), respectively, while the dotted line is the upper bound obtained from (18) when $|\tilde{d}_1|$ is replaced by the bound in (22). Again, the estimate (22) becomes more accurate as harmonic Ritz values converge. The condition number of B is now 707.4, resulting in less of a difference between the computed value of $|\tilde{d}_1|\,\|B\|$ on which

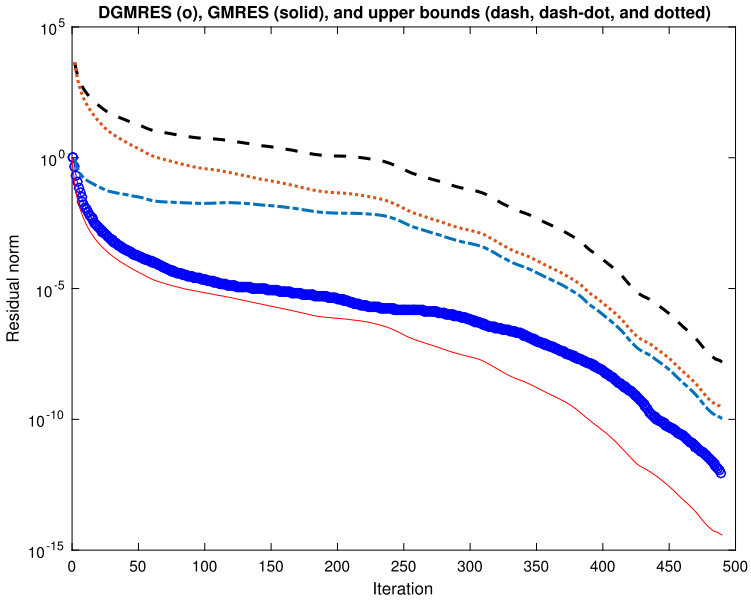

Fig. 1. Convergence of DGMRES for a matrix of index 1 arising from a Neumann problem for Poisson's equation (○'s) and GMRES applied to the nonsingular part (solid). Upper bounds (18) (dash–dot) and (20) (dashed) and (18) with $|\tilde{d}_1|$ replaced by (22) (dotted).

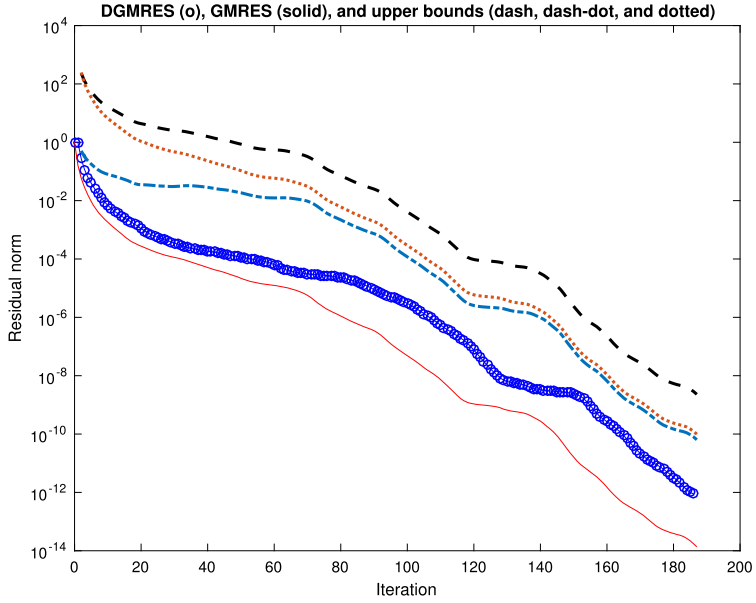

Fig. 2. Convergence of DGMRES for the matrix of Fig. 1 with an incomplete Cholesky preconditioner (○'s) and GMRES applied to the nonsingular part (solid). Upper bounds (18) (dash–dot) and (20) (dashed) and (18) with $|\tilde{d}_1|$ replaced by (22) (dotted).

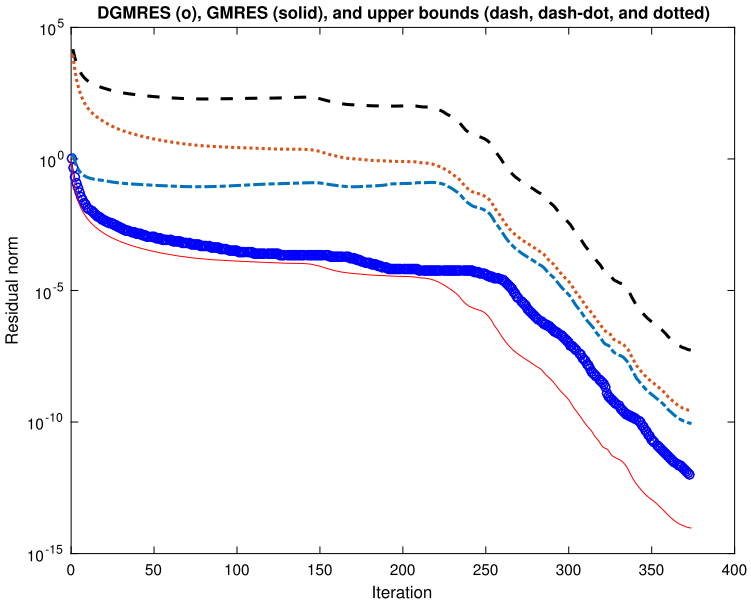

Fig. 3. Convergence of DGMRES for the finite difference matrix arising from equation (35) (∘'s) and GMRES applied to the nonsingular part (solid). Upper bounds (18) (dash–dot) and (20) (dashed) and (18) with $|\tilde{d}_1|$ replaced by (22) (dotted).

the dash–dot curve is based, and the upper bound $|\tilde{d}_1|\ \|B\| \leq (k+1)\kappa(B)$, on which the dashed curve is based.

4.2. Nonsymmetric problems with index 1

To illustrate the use of DGMRES and the residual norm bounds for a nonsymmetric problem with index 1, we simply added some first derivative terms to the Neumann problem of the previous subsection:

$$-(u_{xx} + u_{yy}) + c_1 u_x + c_2 u_y = f \text{ in } [0,1] \times [0,1], \tag{35}$$

again with homogeneous Neumann boundary conditions. Using centered differences for the first derivatives, the finite difference matrix is no longer symmetric or normal, but it still has zero row sums. We set $c_1 = 40$ and $c_2 = -10$, and the results are shown in Fig. 3.

We also set up a problem with multiple zero eigenvalues, but again with index 1. We set a random n by n matrix V of eigenvectors and a diagonal matrix Λ with four zero eigenvalues and the remaining eigenvalues chosen randomly with real parts between -1 and -2 and with imaginary parts taken from a normal distribution with mean 0 and variance 1. Results for $n = 1000$ are shown in Fig. 4. Based on the figure, the comparison between DGMRES and GMRES applied to the nonsingular part seems similar to the previous examples. The upper bound (20) depicted with the dashed curve is a significant

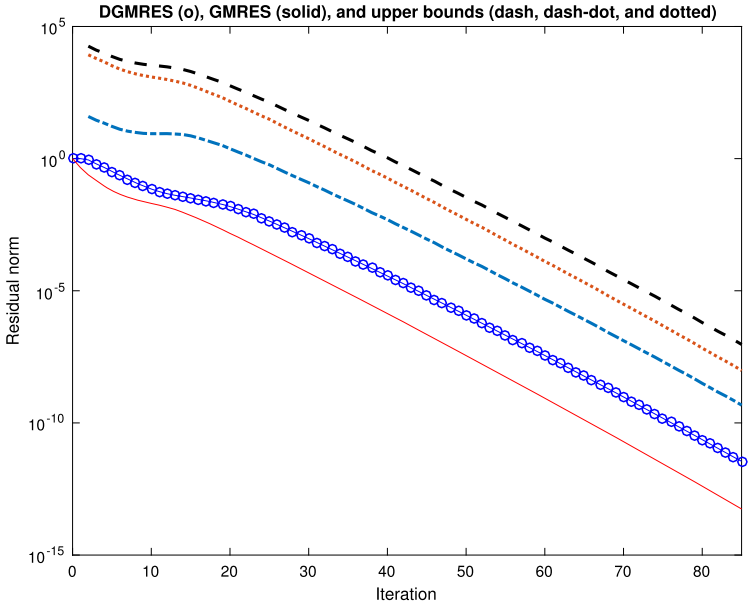

Fig. 4. Convergence of DGMRES for a matrix with random eigenvectors and with eigenvalues chosen randomly with real parts between −1 and −2 and imaginary parts taken from a normal distribution with mean 0 and variance 1 (∘'s). Convergence of GMRES applied to the nonsingular part (solid). Upper bounds (18) (dash–dot) and (20) (dashed) and (18) with $|\tilde{d}_1|$ replaced by (22) (dotted).

overestimate because in this case $\kappa(B) = 2.2e+6$. The bound (18) using the actual value of $|\tilde{d}_1|$ provides a better estimate but is still a few orders of magnitude above the actual DGMRES convergence curve.

4.3. Computing the Drazin inverse

Although one frequently wishes to compute the product $A^D b$ of the Drazin inverse with a given vector b, one sometimes needs to know the entire matrix A^D. One way to compute A^D is to apply DGMRES (with a 0 initial guess) to linear systems whose right-hand sides are each of the n unit vectors. The resulting solution vectors will form the columns of the Drazin inverse. As a simple example, consider the matrix

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 1 & -1 & 0 & 0 \\ -1 & -1 & -1 & 1 & 0 & 0 \\ -1 & -1 & -1 & 0 & 2 & -1 \\ -1 & -1 & 0 & -1 & -1 & 2 \end{bmatrix}$$

which was studied in [2]. The index of this matrix is 2 and its exact Drazin inverse is

$$A^D = \begin{bmatrix} 1/4 & -1/4 & 0 & 0 & 0 & 0 \\ -1/4 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & -1/4 & 0 & 0 \\ 0 & 0 & -1/4 & 1/4 & 0 & 0 \\ 0 & 0 & -5/12 & -7/12 & 2/3 & 1/3 \\ 0 & 0 & -7/12 & -5/12 & 1/3 & 2/3 \end{bmatrix}.$$

Using DGMRES with the unit vectors as right-hand sides, we were able to compute A^D with a relative error in Frobenius norm of $1.3e-15$, which is near the machine precision.

5. Summary and related problems

We have shown that the convergence of DGMRES to the Drazin inverse solution of a possibly inconsistent linear system $Ax = b$ (i.e., to a solution of $A^{\alpha+1}x = A^\alpha b$, where α is greater than or equal to the index of A) is closely related to that of GMRES applied to the nonsingular matrix B in (5). The difference is that the terms $B\hat{r}_0, \ldots, B^\alpha \hat{r}_0$ (or $B, \ldots, B^\alpha$) that are present in the GMRES minimization problems (11)–(13) are not present in the DGMRES minimization problems (8)–(10). An interesting question, independent of the study of DGMRES, is how important are these first terms in the GMRES approximation problem.

Suppose, for instance, that B is Hermitian and positive definite (as it is for the Neumann problem of Section 4.1) with eigenvalues densely distributed in an interval $[a, b]$, so that the worst-case or ideal GMRES polynomial is essentially the polynomial with value one at the origin that minimizes the maximum deviation from 0 on $[a, b]$. This polynomial has the equioscillation property, taking on its maximum absolute value with alternate signs at each of $k+1$ points in the interval if the degree of the polynomial is k. The first degree polynomial is the one that satisfies $p_1^G(a) = -p_1^G(b)$; i.e., $p_1^G(z) = 1 - \frac{2z}{a+b}$. The worst-case or ideal GMRES residual norm reduction at step one is

$$\max_{z\in[a,b]} |p_1^G(z)| = \frac{b-a}{b+a} = \frac{\kappa - 1}{\kappa + 1}, \quad \kappa = \frac{b}{a}.$$

When the DGMRES algorithm is applied to a problem with index α and nonsingular part B, it also constructs a polynomial that equioscillates on the interval $[a, b]$. The first polynomial (of the form $1 - cz^{\alpha+1}$) is determined by requiring that $p_{\alpha+1}^D(a) = -p_{\alpha+1}^D(b)$; that is, $p_{\alpha+1}^D(z) = 1 - \frac{2z^{\alpha+1}}{a^{\alpha+1}+b^{\alpha+1}}$. The worst-case or ideal DGMRES reduction in A^α times the residual at step $\alpha + 1$ is

$$\max_{z\in[a,b]} |p_{\alpha+1}^D(z)| = \frac{b^{\alpha+1} - a^{\alpha+1}}{b^{\alpha+1} + a^{\alpha+1}} = \frac{\kappa^{\alpha+1} - 1}{\kappa^{\alpha+1} + 1}.$$

For κ large, this means *much* slower convergence for DGMRES at the first step where x_0 is modified. [This is not seen in Figs. 1 and 2 of Section 4.1 because the random initial residual is not close to the worst-case initial residual.]

When B is Hermitian, the ideal GMRES or DGMRES polynomials (i.e., the minimax polynomials on the set of eigenvalues of B) can be computed using the Remez algorithm [12]. For an interval $[a, b]$, the GMRES polynomials are shifted and scaled Chebyshev polynomials, leading to the familiar bound

$$\frac{\|r_k^G\|}{\|r_0^G\|} \leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k .$$

Beyond $k = \alpha+1$, however, we know of no analytic estimates for the size of the polynomial of the form $p_k(z) = 1 - c_1 z^{\alpha+1} - \ldots - c_{k-\alpha} z^k$ whose maximum deviation from zero on $[a, b]$ is as small as possible. This could be an interesting question, independent of its connection to DGMRES. Of course, in certain cases the GMRES and DGMRES polynomials are the same. For example, if the matrix B is Hermitian indefinite and has eigenvalues densely distributed throughout two intervals $[-b, -a] \cup [a, b]$ symmetric about the origin, then the ideal GMRES polynomial involves only even powers of B. Hence if $\alpha \leq 1$ in the DGMRES algorithm, then the DGMRES polynomial will be the same as that of GMRES.

Throughout this paper, we have estimated the importance of the terms $B\hat{r}_0, \ldots, B^\alpha \hat{r}_0$ (or $B, \ldots, B^\alpha$) by first finding the linear combination of all the terms $B\hat{r}_0, \ldots, B^\alpha \hat{r}_0$, $B^{\alpha+1}\hat{r}_0, \ldots, B^k \hat{r}_0$ (or $B, \ldots, B^\alpha, B^{\alpha+1}, \ldots, B^k$) that best approximates $\hat{r}_0$ (or I) and then asking how well the first α terms in that linear combination can be approximated by a linear combination of the remaining terms. A problem with this is that those first α terms may have large norm. For example, in the Neumann problem of Section 4.1, we were surprised to see that the two upper bounds (18) and (20) depicted in Figs. 1 and 2 got closer together rather than further apart as k increased, since the larger bound depended explicitly on k while the smaller one depended only on the coefficient $|\tilde{d}_1|$ of B in the GMRES polynomial. This coefficient (times $\|B\|$) increased with k at a faster rate than $k\, \kappa(B)$. Another approach is to first construct the optimal linear combination of the terms $B^{\alpha+1}\hat{r}_0, \ldots, B^k \hat{r}_0$ (or $B^{\alpha+1}, \ldots, B^k$) and then ask how much better one can do by adding in the lower degree terms. This leads to an estimate in the other direction:

$$\|r_k^G\| \leq \min_{d_1, \ldots, d_\alpha} \|(I - \sum_{j=1}^{\alpha} d_j B^j)\| \, \|A^\alpha r_k^D\|.$$

If B is Hermitian and positive definite, for instance, with condition number κ, then we have the bound

$$\|r_k^G\| \leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^\alpha \|A^\alpha r_k^D\|.$$

Acknowledgements

This paper was written while the second and third authors were spending their sabbatical leaves at the University of Washington, Applied Math Dept. They thank the staff and faculty members of the department for their kind support.

References

[1] R. Bhatia, Matrix Analysis, Springer-Verlag, New York, 1997.
[2] J. Climent, M. Neumann, A. Sidi, A semi-iterative method for real spectrum singular linear systems with an arbitrary index, J. Comput. Appl. Math. 87 (1997) 21–38.
[3] R. Carden, M. Embree, Ritz value localization for non-Hermitian matrices, SIAM J. Matrix Anal. Appl. 33 (2012) 1320–1338.
[4] V. Faber, W. Joubert, M. Knill, T. Manteuffel, Minimal residual method stronger than polynomial preconditioning, SIAM J. Matrix Anal. Appl. 17 (1996) 707–729.
[5] S. Goossens, D. Roose, Ritz and harmonic Ritz values and the convergence of FOM and GMRES, Numer. Linear Algebra Appl. 6 (1999) 281–293.
[6] A. Greenbaum, L.N. Trefethen, GMRES/CR and Arnoldi/Lanczos as matrix approximation problems, SIAM J. Sci. Comput. 15 (1994) 359–368.
[7] N. Guglielmi, M. Overton, G.W. Stewart, An efficient algorithm for computing the generalized null space decomposition, SIAM J. Matrix Anal. Appl. 36 (2015) 38–54.
[8] R. Horn, C. Johnson, Matrix Analysis, second edition, Cambridge, 2013.
[9] J. Liesen, P. Tichý, On best approximation of polynomials in matrices in the matrix 2-norm, SIAM J. Matrix Anal. Appl. 31 (2009) 853–863.
[10] G. Meurant, The coefficients of the FOM and GMRES residual polynomials, SIAM J. Matrix Anal. Appl. 38 (2017) 96–117.
[11] B. Ming, DGMRES method augmented with eigenvectors for computing the Drazin-inverse solution of singular linear systems, Acta Math. Appl. Sin. 32 (2016) 549–558.
[12] E.Ya. Remez, Sur la détermination des polynômes d'approximation de degré donnée, Comm. Soc. Math. Kharkov 10 (1934) 41;
E.Ya. Remez, Sur le calcul effectif des polynomes d'approximation de Tchebychef, C. R. Acad. Sci. 199 (1934) 337–340;
E.Ya. Remez, Sur un procédé convergent d'approximations successives pour déterminer les polynômes d'approximation, C. R. Acad. Sci. 198 (1934) 2063–2065.
[13] A. Sidi, A unified approach to Krylov subspace methods for the Drazin-inverse solution of singular nonsymmetric linear systems, Linear Algebra Appl. 298 (1999) 99–113.
[14] A. Sidi, DGMRES: a GMRES-type algorithm for Drazin-inverse solution of singular nonsymmetric linear systems, Linear Algebra Appl. 335 (2001) 189–204.
[15] G. Sleijpen, H. van der Vorst, A Jacobi–Davidson iterative method for linear eigenvalue problems, SIAM J. Matrix Anal. Appl. 17 (1996) 401–425.
[16] K. Toh, GMRES vs. ideal GMRES, AIAM J. Matrix Anal. Appl. 18 (1997) 30–36.
[17] K.C. Toh, Matrix Approximation Problems and Nonsymmetric Iterative Methods, PhD dissertation, Cornell University, Ithaca, NY, 1996.
[18] F. Toutounian, R. Buzhabadi, New methods for computing the Drazin-inverse solution of singular linear systems, Appl. Math. Comput. 294 (2017) 343–352.