

PROPERTIES OF WORST-CASE GMRES

VANCE FABER*, JÖRG LIESEN† AND PETR TICHÝ‡

Abstract. In the convergence analysis of the GMRES method for a given matrix A , one quantity of interest is the largest possible residual norm that can be attained, at a given iteration step k , over all unit norm initial vectors. This quantity is called the worst-case GMRES residual norm for A and k . We show that the worst case behavior of GMRES for the matrices A and A^T is the same, and we analyze properties of initial vectors for which the worst-case residual norm is attained. In particular, we prove that such vectors satisfy a certain “cross equality”. We show that the worst-case GMRES polynomial may not be uniquely determined, and we consider the relation between the worst-case and the ideal GMRES approximations, giving new examples in which the inequality between the two quantities is strict at all iteration steps $k \geq 3$. Finally, we give a complete characterization of how the values of the approximation problems change in the context of worst-case and ideal GMRES for a real matrix, when one considers complex (rather than real) polynomials and initial vectors.

Key words. GMRES method, worst-case convergence, ideal GMRES, matrix approximation problems, minmax

AMS subject classifications. 65F10, 49K35, 41A52

1. Introduction. Let a nonsingular matrix $A \in \mathbb{R}^{n \times n}$ and a vector $b \in \mathbb{R}^n$ be given. Consider solving the system of linear algebraic equations $Ax = b$ with the initial guess $x_0 = 0$ using the GMRES method. This method generates a sequence of iterates $x_k \in \mathcal{K}_k(A, b) \equiv \text{span}\{b, Ab, \dots, A^{k-1}b\}$, $k = 1, 2, \dots$, so that the corresponding k th residual $r_k \equiv b - Ax_k$ satisfies

$$(1.1) \quad \|r_k\| = \min_{p \in \pi_k} \|p(A)b\|.$$

Here $\|\cdot\|$ denotes the Euclidean norm, and π_k denotes the set of real polynomials of degree at most k and with value one at the origin; see the original paper of Saad and Schultz [14] or, e.g., the books [4, 11, 13].

The convergence analysis of GMRES deals with bounding or estimating the right hand side of (1.1). This is a notoriously difficult problem; see, e.g., the respective chapters in [4, 11, 13]. One way to simplify this problem is to split off the right hand side vector b and to bound or estimate the value of the remaining polynomial matrix approximation problem only, i.e., to consider

$$(1.2) \quad \|r_k\| \leq \varphi_k(A) \|b\|, \quad \text{where} \quad \varphi_k(A) \equiv \min_{p \in \pi_k} \|p(A)\|.$$

Greenbaum and Trefethen nicely described the motivation for this approach in [6, pp. 361–362]. They called $\varphi_k(A)$ the *ideal GMRES* value for A and k , and the (uniquely determined) polynomial that attains this value the *ideal GMRES polynomial* for A and k (see [6, 12] for uniqueness proofs).

*Vanco Research, Big Pine Key, FL 33043, USA (vance.faber@gmail.com).

†Institute of Mathematics, Technical University of Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany (liesen@math.tu-berlin.de). The work of this author was supported by the Heisenberg Program of the Deutsche Forschungsgemeinschaft (DFG).

‡Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 18207 Prague, Czech Republic (tichy@cs.cas.cz). This work was supported by the Grant Agency of the Czech Republic under grant No. P201/13-06684 S, and by the project M100301201 of the institutional support of the Academy of Sciences of the Czech Republic.

Since the majority of the existing GMRES convergence results are (upper or lower) bounds on the ideal GMRES value $\varphi_k(A)$, it is natural to ask how far this value can be from an actual k th residual norm produced by GMRES. This question was formulated by Greenbaum and Trefethen in [6, p. 366], and it can be approached by looking at the following sequence of inequalities that holds for any given $A \in \mathbb{R}^{n \times n}$, integer $k \geq 1$ and unit norm vector $b \in \mathbb{R}^n$:

$$\begin{aligned}
 (1.3) \quad \|r_k\| &= \min_{p \in \pi_k} \|p(A)b\| \\
 &\leq \max_{\|v\|=1} \min_{p \in \pi_k} \|p(A)v\| \equiv \psi_k(A) \\
 &\leq \min_{p \in \pi_k} \max_{\|v\|=1} \|p(A)v\| = \varphi_k(A).
 \end{aligned}$$

The value $\psi_k(A)$ introduced in (1.3) is called the *worst-case GMRES* residual norm for the given A and k . It gives an *attainable upper bound* on all possible k th GMRES residual norms for the given matrix A . A unit norm initial vector and a corresponding polynomial for which the value $\psi_k(A)$ is attained are called a *worst-case GMRES initial vector* and a *worst-case GMRES polynomial* for A and k , respectively.

Let us briefly summarize the most important previous results on worst-case and ideal GMRES (see [15, Sections 1–2] for a more detailed summary). First of all, if A is singular, then $\psi_k(A) = \varphi_k(A) = 1$ for all $k \geq 1$ (to see this, simply take v as a unit norm vector in the kernel of A). Hence only nonsingular matrices A are of interest in our context. For such A , both $\psi_k(A)$ and $\varphi_k(A)$ are monotonically decreasing sequences, and $\psi_k(A) = \varphi_k(A) = 0$ for all $k \geq d(A)$, the degree of the minimal polynomial of A . Therefore, we only need to consider $1 \leq k \leq d(A) - 1$.

For a fixed k , both $\psi_k(A)$ and $\varphi_k(A)$ are continuous functions on the open set of nonsingular matrices; see [7, Theorem 3.1] or [2, Theorem 2.5]. Moreover, the equality $\psi_k(A) = \varphi_k(A)$ holds for normal matrices A and any k , as well as for $k = 1$ and any nonsingular A [5, 8]. Some nonnormal matrices A are known, however, for which $\psi_k(A) < \varphi_k(A)$, even $\psi_k(A) \ll \varphi_k(A)$, for certain k ; see [2, 16].

As shown in [18], the ideal GMRES approximation problem can be formulated as a semidefinite program. Hence the ideal GMRES value $\varphi_k(A)$ and the corresponding ideal GMRES polynomial can be computed by any suitably applied semidefinite program solver. In our computations we use the MATLAB package SDPT3, version 4.0; see, e.g., [17]. On the other hand, we are not aware of any efficient algorithm for solving the worst-case GMRES approximation problem. In our experiments we use the general purpose nonlinear minimization routine `fminsearch` from MATLAB's Optimization Toolbox.

Our main goal in this paper is to contribute to the understanding of the worst-case GMRES approximation problem (1.3). In particular, we will derive special properties of worst-case GMRES initial vectors, and we will show that (in contrast to ideal GMRES), worst-case GMRES polynomials for given A and k may not be uniquely determined. Furthermore, we will give some new results on the relation between worst-case and ideal GMRES, and on the tightness of the inequality $\psi_k(A) \leq \varphi_k(A)$. Finally, we give a complete characterization of how the values of the approximation problems in the context of worst-case and ideal GMRES for a real matrix change, when one considers complex (rather than real) polynomials and initial vectors.

In this paper we do not consider quantitative estimation of the worst-case GMRES residual norm $\psi_k(A)$, and we do not study how this value depends on properties of A . This is an important problem of great practical interest, which is largely open. For more details and a survey of the current state-of-the-art we refer to [11, Section 5.7].

2. The cross equality. In this section we generalize two results of Zavorin [19]. The first shows that $\psi_k(A) = \psi_k(A^T)$ and the second is a special property of worst-case initial vectors (they satisfy the so-called ‘‘cross equality’’). Zavorin proved these results only for diagonalizable matrices using quite a complicated technique based on a decomposition of the corresponding Krylov matrix. Using a simple algebraic technique we prove these results for general matrices.

In our derivation we will use the following notation and basic facts about GMRES. For any given nonsingular $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$ the sequence of GMRES residual norms $\|r_k\|$, $k = 1, 2, \dots$, is monotonically decreasing. It terminates with $r_k = 0$ if and only if k is equal to $d(A, b)$, the degree of the minimal polynomial of b with respect to A , where always $d(A, b) \leq d(A)$. A geometric characterization of the GMRES iterate $x_k \in \mathcal{K}_k(A, b)$, which is mathematically equivalent to (1.1), is given by

$$(2.1) \quad r_k \perp AK_k(A, b).$$

When we need to emphasize the dependence of the k th GMRES residual r_k on A , b and k we will write

$$r_k = \text{GMRES}(A, b, k) \quad \text{or} \quad r_k = p_k(A)b,$$

where $p_k \in \pi_k$ is the k th GMRES polynomial of A and b , i.e., the polynomial that solves the minimization problem on the right hand side of (1.1). As long as $r_k \neq 0$, this polynomial is uniquely determined.

LEMMA 2.1. *Let $A \in \mathbb{R}^{n \times n}$ be nonsingular, $k \geq 1$, and let $b \in \mathbb{R}^n$ be a unit norm vector such that $d(A, b) > k$. Let*

$$r_k = \text{GMRES}(A, b, k), \quad s_k = \text{GMRES}\left(A^T, \frac{r_k}{\|r_k\|}, k\right).$$

Then

$$(2.2) \quad \|r_k\| \leq \|s_k\|$$

with equality if and only if

$$\frac{s_k}{\|s_k\|} = b.$$

As a consequence, if $d(A, b) > k$, then also $d(A^T, r_k) > k$.

Proof. Consider any unit norm vector b such that $1 \leq k < d(A, b)$. Then the corresponding k th GMRES residual vector $r_k = p_k(A)b$ is nonzero. The defining property (2.1) of r_k means that $\langle A^j b, r_k \rangle = 0$ for $j = 1, \dots, k$. Hence, for any $q \in \pi_k$,

$$(2.3) \quad \|r_k\|^2 = \langle p_k(A)b, r_k \rangle = \langle b, r_k \rangle = \langle q(A)b, r_k \rangle = \langle b, q(A^T)r_k \rangle \leq \|q(A^T)r_k\|,$$

where the inequality follows from the Cauchy-Schwarz inequality and $\|b\| = 1$. Taking the minimum over all $q \in \pi_k$ in (2.3) and dividing by $\|r_k\|$ we get

$$\|r_k\| \leq \min_{q \in \pi_k} \left\| q(A^T) \frac{r_k}{\|r_k\|} \right\| = \|s_k\|.$$

Now $\|r_k\| > 0$ implies $\|s_k\| > 0$ and hence $d(A^T, r_k) > k$.

Next consider $s_k = q_k(A^T) \frac{r_k}{\|r_k\|}$ and substitute q_k for q into (2.3) to obtain

$$(2.4) \quad \|r_k\|^2 = \langle b, q_k(A^T)r_k \rangle \leq \|q_k(A^T)r_k\| = \|r_k\| \|s_k\|.$$

Therefore, $\|r_k\| = \|s_k\|$ if and only if

$$\langle b, q_k(A^T)r_k \rangle = \|q_k(A^T)r_k\|.$$

Since $\|b\| = 1$, this happens if and only if

$$b = \frac{q_k(A^T)r_k}{\|q_k(A^T)r_k\|} = \frac{q_k(A^T)r_k}{\|s_k\|\|r_k\|} = \frac{s_k}{\|s_k\|},$$

which finishes the proof. \square

We now can show that the worst-case GMRES residual norms for A and A^T are identical.

THEOREM 2.2. *If $A \in \mathbb{R}^{n \times n}$ is nonsingular, then $\psi_k(A) = \psi_k(A^T)$ for all $k = 1, \dots, d(A) - 1$.*

Proof. If b is a worst-case GMRES initial vector for A and k , $r_k = \text{GMRES}(A, b, k)$ and $s_k = \text{GMRES}(A^T, \frac{r_k}{\|r_k\|}, k)$, then, using Lemma 2.1,

$$(2.5) \quad \psi_k(A) = \|r_k\| \leq \|s_k\| \leq \psi_k(A^T).$$

Now we can reverse the roles of A and A^T to obtain the opposite inequality, i.e., $\psi_k(A^T) \leq \psi_k(A)$. \square

The following theorem describes a special property of worst-case initial vectors.

THEOREM 2.3. *Let $A \in \mathbb{R}^{n \times n}$ be nonsingular, and let $1 \leq k \leq d(A) - 1$. If $b \in \mathbb{R}^n$ is a worst-case GMRES initial vector for A and k , and*

$$\begin{aligned} r_k &= p_k(A)b = \text{GMRES}(A, b, k), \\ s_k &= q_k(A^T) \frac{r_k}{\|r_k\|} = \text{GMRES}\left(A^T, \frac{r_k}{\|r_k\|}, k\right), \end{aligned}$$

then

$$\|s_k\| = \|r_k\| = \psi_k(A), \quad b = \frac{s_k}{\psi_k(A)},$$

and

$$(2.6) \quad q_k(A^T)p_k(A)b = \psi_k^2(A)b.$$

Proof. By assumption, $\|r_k\| = \psi_k(A)$. Using Lemma 2.1 and Theorem 2.2,

$$\psi_k(A^T) = \psi_k(A) = \|r_k\| \leq \|s_k\| \leq \psi_k(A^T).$$

Therefore, $\|r_k\| = \|s_k\| = \psi_k(A)$. Using Lemma 2.1 we obtain

$$b = \frac{s_k}{\|s_k\|} = \frac{s_k}{\psi_k(A)},$$

so that $q_k(A^T)p_k(A)b = q_k(A^T)r_k = \|r_k\|s_k = \psi_k^2(A)b$. \square

Equation (2.6) shows that b is an eigenvector of the matrix $q_k(A^T)p_k(A)$ with the corresponding eigenvalue $\psi_k^2(A)$. In Corollary 3.7 we will show that $q_k = p_k$, i.e., that b is a right singular vector of the matrix $p_k(A)$.

To further investigate vectors with the special property introduced in Theorem 2.3 we use the following definition.

DEFINITION 2.4. *Let $A \in \mathbb{R}^{n \times n}$ be nonsingular and $k \geq 1$. We say that a unit norm vector $b \in \mathbb{R}^n$ with $d(A, b) > k$ satisfies the cross equality for A and k , if*

$$b = \frac{s_k}{\|s_k\|}, \quad \text{where } s_k \equiv \text{GMRES}\left(A^T, \frac{r_k}{\|r_k\|}, k\right), \quad r_k \equiv \text{GMRES}(A, b, k).$$

The following algorithm is motivated by this definition. Convergence properties are shown in the theorem immediately below the algorithm statement.

Algorithm 1 (Cross iterations 1)

```

 $b^{(0)} = b,$ 
for  $j = 1, 2, \dots$  do
   $r_k^{(j)} = \text{GMRES}(A, b^{(j-1)}, k)$ 
   $c^{(j-1)} = r_k^{(j)} / \|r_k^{(j)}\|$ 
   $s_k^{(j)} = \text{GMRES}(A^T, c^{(j-1)}, k)$ 
   $b^{(j)} = s_k^{(j)} / \|s_k^{(j)}\|$ 
end for

```

THEOREM 2.5. *Let $A \in \mathbb{R}^{n \times n}$ be nonsingular and $k \geq 1$. If $b \in \mathbb{R}^n$ is any unit norm vector with $d(A, b) > k$, then the vectors generated by Algorithm 1 are well defined and it holds that*

$$(2.7) \quad \|r_k^{(j)}\| \leq \|s_k^{(j)}\| \leq \|r_k^{(j+1)}\| \leq \|s_k^{(j+1)}\| \leq \psi_k(A), \quad j = 1, 2, \dots,$$

and the two sequences $\|r_k^{(j)}\|$, $j = 1, 2, \dots$, and $\|s_k^{(j)}\|$, $j = 1, 2, \dots$, converge to the same limit. Moreover,

$$\lim_{j \rightarrow \infty} \|b^{(j)} - b^{(j-1)}\| = 0 \quad \text{and} \quad \lim_{j \rightarrow \infty} \|c^{(j)} - c^{(j-1)}\| = 0.$$

Proof. Using Lemma 2.1 we know that $r_k^{(1)}$ as well as $s_k^{(1)}$ are well defined and it holds that $\|r_k^{(1)}\| \leq \|s_k^{(1)}\|$. Switching the roles of A and A^T and using Lemma 2.1 again, it follows that $r_k^{(2)}$ is well defined and that $\|s_k^{(1)}\| \leq \|r_k^{(2)}\|$. Hence, (2.7) follows from Lemma 2.1 by induction.

By (2.7) the two sequences $\|r_k^{(j)}\|$ and $\|s_k^{(j)}\|$ interlace each other, are both nondecreasing, and are both bounded by $\psi_k(A)$. This implies that both sequences converge to the same limit, which does not exceed $\psi_k(A)$.

The first equality in (2.4) shows that $\|r_k^{(j)}\| = \langle b^{(j-1)}, s_k^{(j)} \rangle$. Using this fact and $b^{(j)} = s_k^{(j)} / \|s_k^{(j)}\|$ we obtain

$$\frac{1}{2} \|b^{(j)} - b^{(j-1)}\|^2 = 1 - \langle b^{(j-1)}, b^{(j)} \rangle = 1 - \langle b^{(j-1)}, s_k^{(j)} / \|s_k^{(j)}\| \rangle = 1 - \frac{\|r_k^{(j)}\|}{\|s_k^{(j)}\|}.$$

Since the sequences of norms $\|r_k^{(j)}\|$ and $\|s_k^{(j)}\|$ converge to the same limit for $j \rightarrow \infty$, their ratio converges to 1, so that $\|b^{(j)} - b^{(j-1)}\| \rightarrow 0$ for $j \rightarrow \infty$.

The proof of the property for the sequence $c^{(j)}$ is analogous. \square

The results in Theorem 2.5 can be interpreted as a generalization of a theorem of Forsythe from 1968 [3, Theorem 3.8] from symmetric positive definite A to general nonsingular A . As already noticed by Forsythe (for the symmetric positive definite case), there is a strong numerical evidence that for each initial $b^{(0)}$ the sequence $b^{(j)}$ (resp. the sequence $c^{(j)}$) converges to a uniquely defined limit vector \tilde{b} (resp. \tilde{c}). Unfortunately, we were not able to prove that this must always be the case. Such proof could be used to settle the conjecture made by Forsythe in [3, p. 66]. For a recent treatment and historical notes on this open problem we refer to [1].

From the above it is clear that satisfying the cross equality represents a necessary condition for a vector $b^{(0)}$ to be a worst-case initial vector. On the other hand, we can ask whether this condition is sufficient, or, at least, whether the vectors that satisfy the cross equality are in some sense special. To investigate this question we present the following two lemmas.

LEMMA 2.6. *Let $A \in \mathbb{R}^{n \times n}$ be nonsingular and $k \geq 1$. A unit norm vector $b \in \mathbb{R}^n$ with $d(A, b) > k$ satisfies the cross equality for A and k if and only if $b \in \mathcal{K}_{k+1}(A^T, r_k)$, where $r_k = \text{GMRES}(A, b, k)$. In particular, if $d(A) = n$, then each unit norm vector b with $d(A, b) = n$ satisfies the cross equality for A and $k = n - 1$.*

Proof. The nonzero GMRES residual $r_k \in b + AK_k(A, b) \subset \mathcal{K}_{k+1}(A, b)$ is uniquely determined by the orthogonality conditions (2.1), which can be written as

$$0 = \langle A^j b, r_k \rangle = \langle b, (A^T)^j r_k \rangle, \quad \text{for } j = 1, \dots, k,$$

or, equivalently,

$$(2.8) \quad b \perp A^T \mathcal{K}_k(A^T, r_k).$$

Now let $s_k = \text{GMRES}(A^T, r_k / \|r_k\|, k)$. Then

$$(2.9) \quad s_k \in \frac{r_k}{\|r_k\|} + A^T \mathcal{K}_k(A^T, r_k) \subset \mathcal{K}_{k+1}(A^T, r_k), \quad s_k \perp A^T \mathcal{K}_k(A^T, r_k).$$

We will now prove the equivalence. On the one hand, if b satisfies the cross equality for A and k , then $b = s_k / \|s_k\|$ and (2.9) implies that $b \in \mathcal{K}_{k+1}(A^T, r_k)$.

On the other hand, suppose that $b \in \mathcal{K}_{k+1}(A^T, r_k)$. From (2.8) it follows that also $b \perp A^T \mathcal{K}_k(A^T, r_k)$. Since $A^T \mathcal{K}_k(A^T, r_k)$ is a k -dimensional subspace of the $(k+1)$ -dimensional subspace $\mathcal{K}_{k+1}(A^T, r_k)$, b has to be a multiple of s_k , i.e. $b = s_k / \|s_k\|$ or $b = -s_k / \|s_k\|$. Finally, from (2.9) we get $\langle b, s_k \rangle = \|r_k\|^{-1} \langle b, r_k \rangle = \|r_k\| > 0$. Therefore, $b = s_k / \|s_k\|$.

For $k = n - 1$, we have $\mathcal{K}_{k+1}(A^T, r_k) = \mathbb{R}^n$, i.e. $b \in \mathcal{K}_{k+1}(A^T, r_k)$ is always satisfied. \square

LEMMA 2.7. *Let*

$$(2.10) \quad J_\lambda = \begin{bmatrix} \lambda & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & 1 & \\ & & & \ddots & \lambda \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad \lambda \neq 0.$$

Then $e_n = [0, \dots, 0, 1]^T$ satisfies the cross equality for J_λ and every $k = 1, \dots, n-1$.

Proof. From [10, Example 2.3] we know that

$$(2.11) \quad r_k = \text{GMRES}(J_\lambda, e_n, k) = \|r_k\|^2 [0, \dots, 0, (-\lambda)^k, (-\lambda)^{k-1}, \dots, -\lambda, 1]^T.$$

Using Lemma 2.6, it is sufficient to show that $e_n \in \mathcal{K}_{k+1}(J_\lambda^T, r_k)$. We will look at the nonzero structure of the vectors $(J_\lambda^T)^j r_k$. First, it holds that

$$J_\lambda^T r_k = (-1)^k \|r_k\|^2 \lambda^{k+1} e_{n-k}.$$

Consequently, for $j = 1, \dots, k-1$, $(J_\lambda^T)^{j+1} r_k = (J_\lambda^T)^j (J_\lambda^T r_k)$ is a nonzero multiple of the $(n-k)$ th column of $(J_\lambda^T)^j$. Hence

$$[r_k, J_\lambda^T r_k, \dots, (J_\lambda^T)^k r_k] = \begin{bmatrix} \circ & \dots & \dots & \circ \\ \vdots & & & \vdots \\ \circ & \dots & \dots & \circ \\ \bullet & \bullet & \dots & \bullet \\ \bullet & \circ & \ddots & \vdots \\ \vdots & \vdots & \ddots & \bullet \\ \bullet & \circ & \dots & \circ \end{bmatrix},$$

where “ \bullet ” stands for a nonzero entry and “ \circ ” represents a zero entry. From this structure one can easily see that $e_n \in \mathcal{K}_{k+1}(J_\lambda^T, r_k)$. \square

Our numerical tests predict that although e_n satisfies the cross equality for J_λ and every $k = 1, \dots, n-1$, e_n is not a worst-case GMRES initial vector for J_λ and any k . We are able to prove this statement only in special cases, for example if $1 \leq k \leq n/2$ and $\lambda > 2$. In this case $\psi_k(J_\lambda) = \lambda^{-k}$ (cf. [15, Corollary 3.3]), while (2.11) shows that $r_k = \text{GMRES}(J_\lambda, e_n, k)$ has the norm

$$\|r_k\| = \left(\lambda^{2k} + \sum_{j=0}^{k-1} \lambda^{2j} \right)^{-1/2} < \lambda^{-k}.$$

To give a numerical example for Algorithm 1 we consider $A = J_1 \in \mathbb{R}^{11 \times 11}$ and $k = 5$. In the left part of Fig. 2.1 we plot the results of Algorithm 1 started with four random unit norm initial vectors and executed for $j = 1, 2, \dots, 10$. Each line represents one corresponding sequence $\|r_5^{(1)}\|, \|s_5^{(1)}\|, \|r_5^{(2)}\|, \|s_5^{(2)}\|, \dots, \|r_5^{(10)}\|, \|s_5^{(10)}\|$. In each case we noted that the sequences numerically converge to uniquely defined limit vectors (cf. our remarks following the proof of Theorem 2.5). Moreover, in each case we obtain at the end a unit norm vector $b^{(10)}$ that satisfies (up to a small inaccuracy) the cross equality for J_1 and $k = 5$. We can observe that there seems to be no special structure in the norms that are attained at the end. In particular, none of the runs results in a worst-case initial vector for J_1 and $k = 5$, i.e., none of the curves attains the value $\psi_5(J_\lambda)$ that is visualized by the highest bold horizontal line in the figure.

As indicated in the left part of Fig. 2.1, the sequences of residual norms generated by Algorithm 1 usually stagnate after only a few iterations. Unfortunately, this level is usually far below the worst-case level we want to reach. In order to get closer to

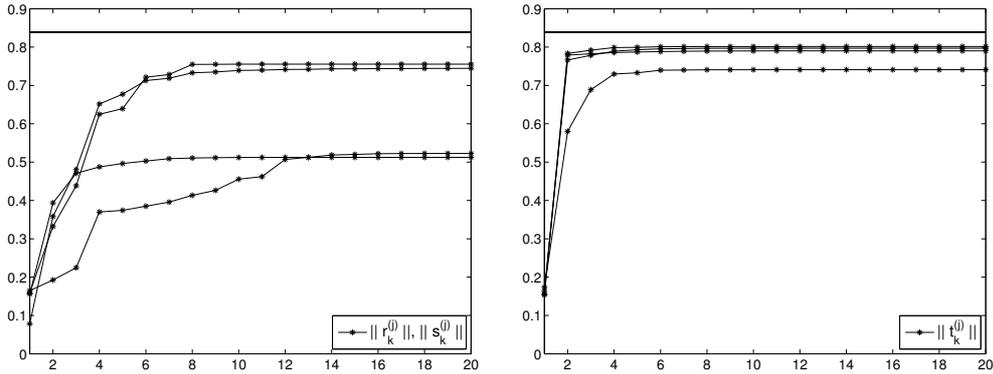


FIG. 2.1. Cross iterations for the 11×11 Jordan block J_1 , $k = 5$, and four different random initial vectors. The left part shows results for Algorithm 1 and the right part for Algorithm 2. The bold solid horizontal line represents the worst-case GMRES residual norm for J_1 .

that level, we need to disturb the process and try a different initial vector that could provide a greater GMRES residual norm. This motivates the following modification of Algorithm 1, where in each step we decide between using A or A^T to generate the next residual norm.

Algorithm 2 (Cross iterations 2)

```

 $b^{(0)} = b,$ 
for  $j = 1, 2, \dots$  do
   $v = \text{GMRES}(A, b^{(j-1)}, k)$ 
   $w = \text{GMRES}(A^T, b^{(j-1)}, k)$ 
  if  $\|v\| < \|w\|$  then
     $t_k^{(j)} = w$ 
  else
     $t_k^{(j)} = v$ 
  end if
   $b^{(j)} = t_k^{(j)} / \|t_k^{(j)}\|$ 
end for

```

Algorithm 2 is well defined and has similar convergence properties as those stated in Theorem 2.5 for Algorithm 1. As shown in the right part of Fig. 2.1, the strategy of Algorithm 2 is a little bit better than the one of Algorithm 1 when looking for a worst-case initial vector: It generates larger residual norms than Algorithm 1, but they are still less than the true worst-case norm. While one may use the output of Algorithm 2 as an initial point for an optimization routine like `fminsearch`, finding an efficient algorithm for computing a worst-case initial vector remains an open problem.

3. Optimization point of view. In this section we rewrite the worst-case GMRES approximation problem (1.3) in an equivalent form in order to characterize worst-case GMRES initial vectors and the corresponding worst-case GMRES polynomials as saddle points of a certain function. This formulation will in particular be used to show that the worst-case GMRES polynomials for A and A^T are identical.

Let a nonsingular matrix $A \in \mathbb{R}^{n \times n}$ and a positive integer $k < d(A)$ be given.

For vectors $c = [c_1, \dots, c_k]^T \in \mathbb{R}^k$ and $v \in \mathbb{R}^n$, we define the function

$$(3.1) \quad f(c, v) \equiv \|p(A; c)v\|^2 = v^T p(A; c)^T p(A; c)v,$$

where

$$p(z; c) = 1 - \sum_{j=1}^k c_j z^j.$$

Equivalently, we can express the function $f(c, v)$ using the matrix

$$K(v) \equiv [Av, A^2v, \dots, A^k v]$$

as

$$(3.2) \quad f(c, v) = \|v - K(v)c\|^2 = v^T v - 2v^T K(v)c + c^T K(v)^T K(v)c.$$

(Here only the dependence on v is expressed in the notation $K(v)$, because A and k are both fixed.) Note that $K(v)^T K(v)$ is the Gramian matrix of the vectors $Av, A^2v, \dots, A^k v$,

$$K(v)^T K(v) = [v^T (A^T)^i A^j v]_{i,j=1,\dots,k}.$$

Next, we define the function

$$g(v) \equiv \min_{c \in \mathbb{R}^k} f(c, v),$$

which represents the k th squared GMRES residual norm for the matrix A and the initial vector v , and we denote

$$\Omega \equiv \{u \in \mathbb{R}^n : d(A, u) \geq k\}, \quad \Gamma \equiv \{u \in \mathbb{R}^n : d(A, u) < k\}.$$

The set Γ is a closed subset, Ω is an open subset of \mathbb{R}^n , and $\mathbb{R}^n = \Omega \cup \Gamma$. Note that $g(v) > 0$ for all $v \in \Omega$ and $g(v) = 0$ for all $v \in \Gamma$. The following lemma is a special case of [2, Proposition 2.2] for real data and nonsingular A .

LEMMA 3.1. *In the previous notation, the function $g(v)$ is a continuous function of $v \in \mathbb{R}^n$, i.e., $g \in C^0(\mathbb{R}^n)$, and it is an infinitely differentiable function of $v \in \Omega$, i.e., $g \in C^\infty(\Omega)$. Moreover, Γ has measure zero in \mathbb{R}^n .*

We next characterize the minimizer of the function $f(c, v)$ as a function of v .

LEMMA 3.2. *For each given $v \in \Omega$, the problem*

$$\min_{c \in \mathbb{R}^k} f(c, v)$$

has the unique minimizer

$$c_*(v) = (K(v)^T K(v))^{-1} K(v)^T v \in \mathbb{R}^k.$$

As a function of $v \in \Omega$, this minimizer satisfies $c_(v) \in C^\infty(\Omega)$. Given $v \in \Omega$, $(c_*(v), v)$ is the only point in $\mathbb{R}^k \times \Omega$ with*

$$\nabla_c f(c_*(v), v) = 0.$$

Proof. Since $v \in \Omega$ and A is nonsingular, the vectors $Av, A^2v, \dots, A^k v$ are linearly independent and $K(v)^T K(v)$ is symmetric and positive definite. Therefore, if $v \in \Omega$ is fixed, (3.2) is a quadratic functional in c , which attains its unique global minimum at the stationary point

$$c_*(v) = (K(v)^T K(v))^{-1} K(v)^T v.$$

Since $K(v)^T K(v)$ is nonsingular and each entry of $(K(v)^T K(v))^{-1}$ can be expressed using Cramer's rule, the function $c_*(v)$ is a well defined rational function of $v \in \Omega$, and thus $c_*(v) \in C^\infty(\Omega)$. Note that the vector $c_*(v)$ contains the coefficients of the k th GMRES polynomial that corresponds to the initial vector $v \in \Omega$. \square

As stated in Lemma 3.1, $g(v)$ is a continuous function on \mathbb{R}^n , and thus it is also continuous on the unit sphere

$$S \equiv \{u \in \mathbb{R}^n : \|u\| = 1\}.$$

Since S is a compact set and $g(v)$ is continuous on this set, it attains its minimum and maximum on S .

We are interested in the characterization of points $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$ such that

$$(3.3) \quad f(\tilde{c}, \tilde{v}) = \max_{v \in S} \min_{c \in \mathbb{R}^k} f(c, v) = \max_{v \in S} g(v).$$

This is the worst-case GMRES problem (1.3). Since $g(v) = 0$ for all $v \in \Gamma$, we have

$$\max_{v \in S} g(v) = \max_{v \in S \cap \Omega} g(v).$$

To characterize the points $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$ that satisfy (3.3), we define for every $c \in \mathbb{R}^k$ and $v \neq 0$ the two functions

$$F(c, v) \equiv f\left(c, \frac{v}{\|v\|}\right) = \frac{f(c, v)}{v^T v}, \quad G(v) \equiv g\left(\frac{v}{\|v\|}\right) = \frac{g(v)}{v^T v}.$$

Clearly, for any $\alpha \neq 0$, we have

$$F(c, \alpha v) = F(c, v), \quad G(\alpha v) = G(v).$$

LEMMA 3.3. *It holds that $G(v) \in C^\infty(\Omega)$. A vector $\tilde{v} \in \Omega \cap S$ satisfies*

$$g(\tilde{v}) \geq g(v) \quad \text{for all } v \in S$$

if and only if $\tilde{v} \in \Omega \cap S$ satisfies

$$G(\tilde{v}) \geq G(v) \quad \text{for all } v \in \mathbb{R}^n \setminus \{0\}.$$

Proof. Since $g(v) \in C^\infty(\Omega)$ and $0 \notin \Omega$, it holds also $G(v) \in C^\infty(\Omega)$. If $\tilde{v} \in \Omega \cap S$ is a maximum of $G(v)$, then $\alpha \tilde{v}$ is a maximum as well, so the equivalence is obvious. \square

THEOREM 3.4. *The vectors $\tilde{c} \in \mathbb{R}^k$ and $\tilde{v} \in S \cap \Omega$ that solve the problem*

$$\max_{v \in S} \min_{c \in \mathbb{R}^k} f(c, v)$$

satisfy

$$(3.4) \quad \nabla_c F(\tilde{c}, \tilde{v}) = 0, \quad \nabla_v F(\tilde{c}, \tilde{v}) = 0,$$

i.e., (\tilde{c}, \tilde{v}) is a stationary point of the function $F(c, v)$.

Proof. Obviously, for any $v \in \Omega$,

$$F(c_*(v), v) = \frac{f(c_*(v), v)}{v^T v} \leq \frac{f(c, v)}{v^T v} = F(c, v) \quad \text{for all } c \in \mathbb{R}^k,$$

i.e., $c_*(v)$ also minimizes the function $F(c, v)$. Hence,

$$\nabla_c F(c_*(v), v) = 0, \quad v \in \Omega.$$

We know that $g(v)$ attains its maximum on S at some point $\tilde{v} \in \Omega \cap S$. Therefore, $G(v)$ attains its maximum also at \tilde{v} . Since $G(v) \in C^\infty(\Omega)$, it has to hold that

$$\nabla G(\tilde{v}) = 0.$$

Denoting $\tilde{c} = c_*(\tilde{v})$ and writing the function $G(v)$ as $G(v) = F(c_*(v), v)$ we get

$$(3.5) \quad \nabla G(\tilde{v}) = 0 = \nabla_v c_*(\tilde{v}) \nabla_c F(\tilde{c}, \tilde{v}) + \nabla_v F(\tilde{c}, \tilde{v}),$$

where $\nabla_v c_*(\tilde{v})$ is the $n \times k$ Jacobian matrix of the function $c_*(v) : \mathbb{R}^n \rightarrow \mathbb{R}^k$ at the point \tilde{v} . Here we used the standard chain rule for multivariate functions. Since $\tilde{v} \in \Omega \cap S$, we know that $\nabla_c F(\tilde{c}, \tilde{v}) = 0$, and, therefore, using (3.5), $\nabla_v F(\tilde{c}, \tilde{v}) = 0$. \square

THEOREM 3.5. *If (\tilde{c}, \tilde{v}) is a solution of the problem (3.3), then \tilde{v} is a right singular vector of the matrix $p(A; \tilde{c})$.*

Proof. Since (\tilde{c}, \tilde{v}) solves the problem (3.3), we have $0 = \nabla_v F(\tilde{c}, \tilde{v})$. Writing $F(c, v)$ as a Rayleigh quotient,

$$F(c, v) = \frac{v^T p(A; c)^T p(A; c) v}{v^T v},$$

we ask when $\nabla_v F(c, v) = 0$; for more details see [9, pp. 114–115]. By differentiating $F(c, v)$ with respect to v we get

$$0 = \frac{2p(A; c)^T p(A; c) v \|v\|^2 - 2[v^T p(A; c)^T p(A; c) v] v}{(v^T v)^2}$$

and the condition $0 = \nabla_v F(\tilde{c}, \tilde{v})$ is equivalent to

$$p(A; \tilde{c})^T p(A; \tilde{c}) \tilde{v} = F(\tilde{c}, \tilde{v}) \tilde{v}.$$

In other words, \tilde{v} is a right singular vector of $p(A; \tilde{c})$ and $\sigma = \sqrt{F(\tilde{c}, \tilde{v})}$ is the corresponding singular value. \square

THEOREM 3.6. *A point $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$ that solves the problem (3.3) is a stationary point of $F(c, v)$ in which the maximal value of $F(c, v)$ is attained.*

Proof. Using Theorem 3.4 we know that any solution $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$ of (3.3) is a stationary point of $F(c, v)$. On the other hand, if $(\hat{c}, \hat{v}) \in \mathbb{R}^k \times S$ satisfies

$$\nabla_v F(\hat{c}, \hat{v}) = 0, \quad \nabla_c F(\hat{c}, \hat{v}) = 0,$$

then $p(A; \hat{c})$ is the GMRES polynomial that corresponds to \hat{v} and

$$F(\hat{c}, \hat{v}) = \|p(A; \hat{c})\hat{v}\|^2 \leq \|p(A; \tilde{c})\tilde{v}\|^2 = F(\tilde{c}, \tilde{v}).$$

Hence, (\tilde{c}, \tilde{v}) is a stationary point of $F(c, v)$ in which the maximal value of $F(c, v)$ is attained. \square

As a consequence of previous results we can formulate the following corollary.

COROLLARY 3.7. *With the same assumptions and notation as in Theorem 2.3, it holds that $p_k = q_k$.*

Proof. Using Theorem 3.5 and Theorem 3.6 we know that

$$(3.6) \quad \psi_k^2(A)b = p_k(A^T)p_k(A)b,$$

i.e., that b is a right singular vector of the matrix $p_k(A)$ that corresponds to the maximal value of $F(\tilde{c}, \tilde{v})$, i.e., to $\psi_k^2(A)$. From (2.6) we also know that

$$(3.7) \quad \psi_k^2(A)b = q_k(A^T)p_k(A)b$$

where q_k is the GMRES polynomial that corresponds to A^T and the initial vector r_k . Comparing (3.6) and (3.7), and using the uniqueness of the GMRES polynomial q_k it follows that $p_k = q_k$. \square

4. Non-uniqueness of worst-case GMRES polynomials. In this section we prove that a worst-case GMRES polynomial may not be uniquely determined, and we give a numerical example for the occurrence of a non-unique case. Our results are based on Toh's parametrized family of (nonsingular) matrices

$$(4.1) \quad A = A(\omega, \varepsilon) = \begin{bmatrix} 1 & \varepsilon & & \\ & -1 & \frac{\omega}{\varepsilon} & \\ & & 1 & \varepsilon \\ & & & -1 \end{bmatrix} \in \mathbb{R}^{4 \times 4}, \quad 0 < \omega < 2, \quad 0 < \varepsilon.$$

Toh used these matrices in [16] to show that $\psi_3(A)/\varphi_3(A) \rightarrow 0$ for $\varepsilon \rightarrow 0$ and each $\omega \in (0, 2)$ [16, Theorem 2.3]. In other words, he proved that the ratio of the worst-case and ideal GMRES approximations can be arbitrarily small.

THEOREM 4.1. *Let A be as in (4.1). If $p_k(z)$ is a worst-case GMRES polynomial for A and k , then $p_k(-z)$ is also a worst-case GMRES polynomial for A and k .*

In particular, $p_3(z) \neq p_3(-z)$, so the worst-case GMRES polynomial for A and $k = 3$ is not uniquely determined.

Proof. Let b be any worst-case initial vector for A and k , and consider the orthogonal similarity transformation

$$A = -QA^TQ^T, \quad Q = \begin{bmatrix} & & & 1 \\ & & -1 & \\ & 1 & & \\ -1 & & & \end{bmatrix}.$$

Then

$$p_k(A)b = Qp_k(-A^T)Q^Tb \quad \text{and} \quad \psi_k(A) = \|p_k(A)b\| = \|p_k(-A^T)w\| = \psi_k(A^T),$$

where $w = Q^T b$. In other words, $p_k(-z)$ is a worst-case GMRES polynomial for A^T and k . Using Corollary 3.7, it is also a worst-case GMRES polynomial for A and k .

Let $p_3(z) \in \pi_3$ be any worst-case GMRES polynomial for A and $k = 3$. To show that $p_3(-z) \neq p_3(z)$ it suffices to show that $p_3(z)$ contains odd powers of z , i.e., that

$$(4.2) \quad p_3(z) \neq 1 - \beta z^2 \quad \text{for any } \beta \in \mathbb{R}.$$

Define the matrix

$$B \equiv \begin{bmatrix} 1 & 0 & \omega & 0 \\ & 1 & 0 & \omega \\ & & 1 & 0 \\ & & & 1 \end{bmatrix} = A^2.$$

From [16, Theorem 2.1] we know that the (uniquely determined) ideal GMRES polynomial for A and $k = 3$ is of the form

$$(4.3) \quad p_*(z) = 1 + (\alpha - 1)z^2, \quad \alpha = \frac{2\omega^2}{4 + \omega^2}.$$

Therefore,

$$\min_{p \in \pi_3} \|p(A)\| = \min_{p \in \pi_1} \max_{\|v\|=1} \|p(B)v\| = \max_{\|v\|=1} \min_{p \in \pi_1} \|p(B)v\|,$$

where the last equality follows from the fact that the ideal and worst-case GMRES approximations are equal for $k = 1$ [8, 5]. If a worst-case polynomial for A and $k = 3$ is of the form $1 - \beta z^2$ for some β , then

$$\psi_3(A) = \max_{\|v\|=1} \min_{p \in \pi_3} \|p(A)v\| = \max_{\|v\|=1} \min_{p \in \pi_1} \|p(B)v\| = \min_{p \in \pi_3} \|p(A)\| = \varphi_3(A).$$

This, however, contradicts the main result by Toh that $\psi_3(A) < \varphi_3(A)$; see [16, Theorem 2.2]. \square

To compute examples of worst-case GMRES polynomials for the Toh matrix (4.1) numerically we chose $\varepsilon = 0.1$ and $\omega = 1$, and we used the function `fminsearch` from MATLAB's Optimization Toolbox. We computed the value

$$\psi_3(A) = 0.4579$$

(we present the numerical results only to 4 digits) with the corresponding third worst-case initial vector

$$b = [-0.6376, 0.0471, 0.2188, 0.7371]^T$$

and the worst-case GMRES polynomial

$$p_3(z) = -0.025z^3 - 0.895z^2 + 0.243z + 1 = \frac{-1}{39.9}(z - 1.181)(z + 0.939)(z + 35.96).$$

One can numerically check that b is the right singular vector of $p_3(A)$ that corresponds to the second maximal singular value of $p_3(A)$. From Theorem 4.1 we know that $q_3(z) \equiv p_3(-z)$ is also a third worst-case GMRES polynomial. One can now find the

corresponding worst-case initial vector leading to the polynomial q_3 using the singular value decomposition (SVD)

$$p_3(A) = USV^T,$$

where the singular values are ordered nonincreasingly on the diagonal of S . We know (by numerical observation) that b is the second column of V . We now compute the SVD of $q_3(A)$, and define the corresponding initial vector as the right singular vector that corresponds to the second maximal singular value of $q_3(A)$. It holds that

$$p_3(A^T) = p_3(A)^T = VSU^T.$$

Since $A^T = -QAQ^T$, we get $Qp_3(-A)Q^T = VSU^T$, or, equivalently,

$$q_3(A) = (Q^T V)S(Q^T U)^T.$$

So, the columns of the matrix $Q^T U$ are right singular vectors of $q_3(A)$ and the vector $Q^T u_2$, where u_2 is the second column of U , is the worst-case initial vector that gives the worst-case GMRES polynomial $q_3(z) = p_3(-z)$.

5. Ideal versus worst-case GMRES. As mentioned above, Toh [16] as well as Faber, Joubert, Knill, and Manteuffel [2] have shown that worst-case GMRES and ideal GMRES are different approximation problems in the sense that there exist matrices A and iteration steps k for which $\psi_k(A) < \varphi_k(A)$. In this section we further study these two approximation problems. We start with a geometrical characterization related to the function $f(c, v)$ from (3.2).

THEOREM 5.1. *Let $A \in \mathbb{R}^{n \times n}$ be a nonsingular matrix and let $1 \leq k \leq d(A) - 1$. The k th ideal and worst-case GMRES approximations are equal, i.e.,*

$$(5.1) \quad \max_{v \in S} \min_{c \in \mathbb{R}^k} f(c, v) = \min_{c \in \mathbb{R}^k} \max_{v \in S} f(c, v),$$

if and only if $f(c, v)$ has a saddle point in $\mathbb{R}^k \times S$.

Proof. If $f(c, v)$ has a saddle point in $\mathbb{R}^k \times S$, then there exist vectors $\tilde{c} \in \mathbb{R}^k$ and $\tilde{v} \in S$ such that

$$f(\tilde{c}, v) \leq f(\tilde{c}, \tilde{v}) \leq f(c, \tilde{v}) \quad \forall c \in \mathbb{R}^k, \forall v \in S.$$

The condition $f(\tilde{c}, v) \leq f(\tilde{c}, \tilde{v})$ for all $v \in S$ implies that \tilde{v} is a maximal right singular vector of the matrix $p(A; \tilde{c})$. If $f(\tilde{c}, \tilde{v}) \leq f(c, \tilde{v})$ for all $c \in \mathbb{R}^k$, then $p(z; \tilde{c})$ is the GMRES polynomial that corresponds to the initial vector \tilde{v} . In other words, if $f(c, v)$ has a saddle point in $\mathbb{R}^k \times S$, then there exist a polynomial $p(z; \tilde{c})$ and a unit norm vector \tilde{v} such that \tilde{v} is a maximal right singular vector of $p(A; \tilde{c})$ and

$$p(A; \tilde{c})\tilde{v} \perp AK_k(A, \tilde{v}).$$

Using [15, Lemma 2.4], the k th ideal and worst-case GMRES approximations are then equal.

On the other hand, if the condition (5.1) is satisfied, then $f(c, v)$ has a saddle point in $\mathbb{R}^k \times S$. \square

In other words, the k th ideal and worst-case GMRES approximations are equal if and only if the points $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$ that solve the worst-case GMRES problem are also the saddle points of $f(c, v)$ in $\mathbb{R}^k \times S$.

We next extend the original construction of Toh [16] to obtain some further numerical examples in which $\psi_k(A) < \varphi_k(A)$. Note that the Toh matrix (4.1) is not diagonalizable. In particular, for $\omega = 1$ we have $A = X\tilde{J}X^{-1}$, where

$$\tilde{J} = \begin{bmatrix} 1 & 1 & & \\ & 1 & & \\ & & -1 & 1 \\ & & & -1 \end{bmatrix}, \quad X = \begin{bmatrix} \epsilon & \epsilon & \epsilon & -\epsilon \\ -2 & -1 & 0 & 1 \\ 0 & -2\epsilon & 0 & 2\epsilon \\ 0 & 4 & 0 & 0 \end{bmatrix}.$$

One can ask whether the phenomenon $\psi_k(A) < \varphi_k(A)$ can also appear for diagonalizable matrices. The answer is yes, since both $\psi_k(A)$ and $\varphi_k(A)$ are continuous functions on the open set of nonsingular matrices; see [2, Theorem 2.5 and Theorem 2.6]. Hence one can slightly perturb the diagonal of the Toh matrix (4.1) in order to obtain a diagonalizable matrix \tilde{A} for which $\psi_k(\tilde{A}) < \varphi_k(\tilde{A})$.

For $\omega = 1$, the Toh matrix is an upper bidiagonal matrix with the alternating diagonal entries 1 and -1 , and the alternating superdiagonal entries ϵ and ϵ^{-1} . One can consider such a matrix for any $n \geq 4$, i.e.,

$$A = \begin{bmatrix} 1 & \epsilon & & & & & & \\ & -1 & \epsilon^{-1} & & & & & \\ & & 1 & \epsilon & & & & \\ & & & \ddots & \ddots & & & \\ & & & & \ddots & \epsilon^{\pm 1} & & \\ & & & & & \pm 1 & & \end{bmatrix} \in \mathbb{R}^{n \times n},$$

and look at the values of $\psi_k(A)$ and $\varphi_k(A)$. If n is even, we found numerically that $\psi_k(A) = \varphi_k(A)$ for $k \neq n-1$ and $\psi_{n-1}(A) < \varphi_{n-1}(A)$. If n is odd, then our numerical experiments showed that $\psi_k(A) = \varphi_k(A)$ for $k \neq n-2$ and $\psi_{n-2}(A) < \varphi_{n-2}(A)$. Hence for all such matrices worst-case and ideal GMRES differ from each other for exactly one k .

Inspired by the Toh matrix, we define the $n \times n$ matrices (for any $n \geq 2$)

$$J_{\lambda, \epsilon} \equiv \begin{bmatrix} \lambda & \epsilon & & & \\ & \ddots & \ddots & & \\ & & \ddots & \epsilon & \\ & & & \lambda & \end{bmatrix}, \quad E_\epsilon \equiv \begin{bmatrix} 0 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 0 \\ \epsilon^{-1} & 0 & \dots & 0 \end{bmatrix}$$

and use them to construct the matrix

$$A = \begin{bmatrix} J_{1, \epsilon} & \omega E_\epsilon \\ & J_{-1, \epsilon} \end{bmatrix} \in \mathbb{R}^{2n \times 2n}, \quad \omega > 0.$$

One can numerically observe that here $\psi_k(A) < \varphi_k(A)$ for all steps $k = 3, \dots, 2n-1$. As an example, we plot in Fig. 5.1 the ideal and worst-case GMRES convergence curves for $n = 4$, i.e., A is an 8×8 matrix, $\omega = 4$ and $\epsilon = 0.1$. Varying the parameter ω will influence the difference between the worst-case and ideal GMRES values in these examples. Decreasing ω leads to a smaller difference, and increasing ω leads to a larger difference for large k , while the two values need not differ for some small k .

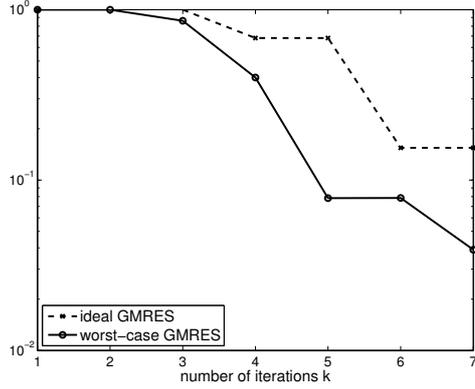


FIG. 5.1. Ideal and worst-case GMRES can differ from step 3 up to the step $2n - 1$.

6. Ideal and worst-case GMRES for complex vectors or polynomials.

We now ask whether the values of the min-max approximation (1.2) and the max-min approximation (1.3) for a matrix $A \in \mathbb{R}^{n \times n}$ can change if we allow the maximization over complex vectors and/or the minimization over complex polynomials. We will give a complete answer to this question in Theorems 6.1 and 6.3 below. In short, for the min-max approximation related to ideal GMRES the underlying fields of minimization and maximization do not matter, while for the max-min approximation related to worst-case GMRES different fields may in some cases indeed lead to different values. These results again indicate the different nature of the two approximation problems, and they complement (and in some sense complete) previous results in the literature dealing with the same question; see, in particular, [2, Section 2],[8, Section 3], and [20, Section 4].

Let us define

$$\varphi_{k,\mathbb{K},\mathbb{F}}(A) \equiv \min_{p \in \pi_{k,\mathbb{K}}} \max_{\substack{b \in \mathbb{F}^n \\ \|b\|=1}} \|p(A)b\|, \quad \psi_{k,\mathbb{K},\mathbb{F}}(A) \equiv \max_{\substack{b \in \mathbb{F}^n \\ \|b\|=1}} \min_{p \in \pi_{k,\mathbb{K}}} \|p(A)b\|,$$

where \mathbb{K} and \mathbb{F} are either the real or the complex numbers. Hence, the previously used $\varphi_k(A)$, $\psi_k(A)$, and π_k are now denoted by $\varphi_{k,\mathbb{R},\mathbb{R}}(A)$ and $\psi_{k,\mathbb{R},\mathbb{R}}(A)$, and $\pi_{k,\mathbb{R}}$, respectively. We first analyze the case of $\varphi_{k,\mathbb{K},\mathbb{F}}(A)$.

THEOREM 6.1. *For a nonsingular matrix $A \in \mathbb{R}^{n \times n}$ and $1 \leq k \leq d(A) - 1$,*

$$\varphi_{k,\mathbb{R},\mathbb{R}}(A) = \varphi_{k,\mathbb{C},\mathbb{R}}(A) = \varphi_{k,\mathbb{R},\mathbb{C}}(A) = \varphi_{k,\mathbb{C},\mathbb{C}}(A).$$

Proof. Since

$$\max_{\substack{b \in \mathbb{R}^n \\ \|b\|=1}} \|Bv\| = \|B\| = \max_{\substack{b \in \mathbb{C}^n \\ \|b\|=1}} \|Bv\|$$

holds for any real matrix $B \in \mathbb{R}^{n \times n}$, we have $\varphi_{k,\mathbb{R},\mathbb{R}}(A) = \varphi_{k,\mathbb{R},\mathbb{C}}(A)$.

Next, from $\mathbb{R} \subset \mathbb{C}$ we get immediately $\varphi_{k,\mathbb{C},\mathbb{R}}(A) \leq \varphi_{k,\mathbb{R},\mathbb{R}}(A)$. On the other hand, writing $p \in \pi_{k,\mathbb{C}}$ in the form $p = p_r + \mathbf{i}p_i$, where $p_r \in \pi_{k,\mathbb{R}}$ and p_i is a real polynomial of degree at most k such that $p_i(0) = 0$, we get

$$\varphi_{k,\mathbb{C},\mathbb{R}}^2(A) = \min_{p \in \pi_{k,\mathbb{C}}} \max_{\substack{b \in \mathbb{R}^n \\ \|b\|=1}} \|p(A)b\|^2 = \min_{p \in \pi_{k,\mathbb{C}}} \max_{\substack{b \in \mathbb{R}^n \\ \|b\|=1}} (\|p_r(A)b\|^2 + \|p_i(A)b\|^2)$$

$$\geq \min_{p_r \in \pi_{k,\mathbb{R}}} \max_{\substack{b \in \mathbb{R}^n \\ \|b\|=1}} \|p_r(A)b\|^2 = \varphi_{k,\mathbb{R},\mathbb{R}}^2(A),$$

so that $\varphi_{k,\mathbb{C},\mathbb{R}}(A) = \varphi_{k,\mathbb{R},\mathbb{R}}(A)$. Finally, from [7, Theorem 3.1] we obtain $\varphi_{k,\mathbb{R},\mathbb{R}}(A) = \varphi_{k,\mathbb{C},\mathbb{C}}(A)$. \square

Since the value of $\varphi_{k,\mathbb{K},\mathbb{F}}(A)$ does not change when choosing for \mathbb{K} and \mathbb{F} real or complex numbers, we will again use the simple notation $\varphi_k(A)$ in the following text. The situation for the quantities corresponding to the worst-case GMRES approximation is more complicated. Our proof of this fact uses the following lemma.

LEMMA 6.2. *If $A = A(\omega, \varepsilon)$ is the Toh matrix defined in (4.1) and*

$$(6.1) \quad B \equiv \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix},$$

then $\psi_{3,\mathbb{R},\mathbb{R}}(B) = \varphi_3(A)$.

Proof. Using the structure of B it is easy to see that $\psi_{k,\mathbb{R},\mathbb{R}}(B) \leq \varphi_k(A)$ for any k . To prove the equality, it suffices to find a real unit norm vector w with

$$(6.2) \quad \min_{p \in \pi_{3,\mathbb{R}}} \|p(B)w\| = \varphi_3(A) = \min_{p \in \pi_{3,\mathbb{R}}} \|p(A)\|.$$

The solution p_* of the ideal GMRES problem on the right hand side of (6.2) is given by (4.3). Toh showed in [16, p. 32] that $p_*(A)$ has a twofold maximal singular value σ , and that the corresponding right and left singular vectors are given (up to a normalization) by

$$[v_1, v_2] = \begin{bmatrix} 0 & \omega \\ \omega & 0 \\ 0 & -2 \\ -2 & 0 \end{bmatrix}, \quad [u_1, u_2] = \begin{bmatrix} 0 & 2 \\ 2 & 0 \\ 0 & -\omega \\ -\omega & 0 \end{bmatrix},$$

i.e., $\sigma u_1 = p_*(A)v_1$ and $\sigma u_2 = p_*(A)v_2$, where $\sigma = \|p_*(A)\|$.

Let us define

$$w \equiv \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} / \left\| \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\|, \quad q(z) \equiv p_*(z).$$

Using

$$q(B) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \sigma \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \text{and} \quad \left\| \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\| = \left\| \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right\|,$$

we see that $\|q(B)w\| = \sigma$. To prove (6.2) it is sufficient to show that q is the third GMRES polynomial for B and w , i.e., that q satisfies $q(B)w \perp B^j w$ for $j = 1, 2, 3$, or, equivalently,

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}^T \begin{bmatrix} A^j & 0 \\ 0 & A^j \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = u_1^T A^j v_1 + u_2^T A^j v_2 = 0, \quad j = 1, 2, 3.$$

Using linear algebra calculations we get $u_1^T A v_1 = -4\omega = -u_2^T A v_2$, and

$$0 = u_1^T A^2 v_1 = u_2^T A^2 v_2 = u_1^T A^3 v_1 = u_2^T A^3 v_2.$$

Therefore, we have found a unit norm initial vector w and the corresponding third GMRES polynomial q such that $\|q(B)w\| = \varphi_3(A)$. \square

We next analyze the quantities $\psi_{k,\mathbb{R},\mathbb{F}}(A)$.

THEOREM 6.3. *For a nonsingular matrix $A \in \mathbb{R}^{n \times n}$ and $1 \leq k \leq d(A) - 1$,*

$$\psi_{k,\mathbb{R},\mathbb{R}}(A) = \psi_{k,\mathbb{C},\mathbb{R}}(A) \leq \psi_{k,\mathbb{C},\mathbb{C}}(A) \leq \psi_{k,\mathbb{R},\mathbb{C}}(A) \leq \varphi_k(A),$$

where the first and the second inequality can be strict.

Proof. For a real initial vector b , the corresponding GMRES polynomial is uniquely determined and real. This implies $\psi_{k,\mathbb{C},\mathbb{R}}(A) = \psi_{k,\mathbb{R},\mathbb{R}}(A)$. Next, from [7, Theorem 3.1] it follows that $\psi_{k,\mathbb{R},\mathbb{R}}(A) \leq \psi_{k,\mathbb{C},\mathbb{C}}(A)$. Finally, using $\mathbb{R} \subset \mathbb{C}$ we get $\psi_{k,\mathbb{C},\mathbb{C}}(A) \leq \psi_{k,\mathbb{R},\mathbb{C}}(A)$. It remains to show that the first and the second inequality can be strict, and that $\psi_{k,\mathbb{R},\mathbb{C}}(A) \leq \varphi_k(A)$.

For the first inequality, as shown in [20, Section 4], there exist real matrices A and certain complex (unit norm) initial vectors b for which $\min_{p \in \pi_{k,\mathbb{C}}} \|p(A)b\| = 1$ for $k = 1, \dots, n - 1$ (complete stagnation), while such complete stagnation does not occur for any real (unit norm) initial vector. Therefore, there are matrices for which $\psi_{k,\mathbb{C},\mathbb{R}}(A) < \psi_{k,\mathbb{C},\mathbb{C}}(A)$.

To show that the second inequality can be strict, we note that for any $A \in \mathbb{R}^{n \times n}$, the corresponding matrix $B \in \mathbb{R}^{2n \times 2n}$ of the form (6.1), and $1 \leq k \leq d(A) - 1$,

$$\begin{aligned} \psi_{k,\mathbb{R},\mathbb{C}}^2(A) &= \max_{\substack{b \in \mathbb{C}^n \\ \|b\|=1}} \min_{p \in \pi_{k,\mathbb{R}}} \|p(A)b\|^2 = \max_{\substack{u, v \in \mathbb{R}^n \\ \|u\|^2 + \|v\|^2 = 1}} \min_{p \in \pi_{k,\mathbb{R}}} \|p(A)(u + \mathbf{i}v)\|^2 \\ &= \max_{\|u\|^2 + \|v\|^2 = 1} \min_{p \in \pi_{k,\mathbb{R}}} (\|p(A)u\|^2 + \|p(A)v\|^2) \\ (6.3) \quad &= \max_{\substack{v \in \mathbb{R}^{2n} \\ \|v\|=1}} \min_{p \in \pi_{k,\mathbb{R}}} \|p(B)v\|^2 = \psi_{k,\mathbb{R},\mathbb{R}}^2(B). \end{aligned}$$

Now let A be the Toh matrix (4.1) and $k = 3$. Toh showed in [16, Theorem 2.2] that for any unit norm $b \in \mathbb{C}^4$ and the corresponding third GMRES polynomial $p_b \in \pi_{3,\mathbb{C}}$,

$$\|p_b(A)b\| < \varphi_3(A).$$

Hence $\psi_{3,\mathbb{C},\mathbb{C}}(A) < \varphi_3(A)$. Lemma 6.2 and equation (6.3) imply $\varphi_3(A) = \psi_{3,\mathbb{R},\mathbb{C}}(A)$ for the Toh matrix, and, therefore, the second inequality can be strict.

Finally, since $\|p(A)\| = \|p(B)\|$ for any polynomial p , we get $\varphi_3(B) = \varphi_3(A)$, and, using (6.3), $\psi_{3,\mathbb{R},\mathbb{C}}(A) = \psi_{3,\mathbb{R},\mathbb{R}}(B) \leq \varphi_3(B) = \varphi_3(A)$. \square

We do not know whether the first and the second inequality in Theorem 6.3 can be strict simultaneously, i.e., can both be strict for the same A and k . Concerning the last inequality in Theorem 6.3, we are in fact able to prove that $\psi_{k,\mathbb{R},\mathbb{C}}(A) = \varphi_k(A)$. Since the techniques used in this proof are beyond the scope of this paper, we will publish it elsewhere.

Our proof concerning the strictness of the first inequality in the previous theorem relied on a numerical example given in [20, Section 4]. We will now give an alternative construction based on the non-uniqueness of the worst-case GMRES polynomial, which will lead to an example with $\psi_{k,\mathbb{R},\mathbb{R}}(A) < \psi_{k,\mathbb{R},\mathbb{C}}(A)$.

Suppose that A is a real matrix for which in a certain step k two *different* worst-case polynomials $p_b \in \pi_{k,\mathbb{R}}$ and $p_c \in \pi_{k,\mathbb{R}}$ with corresponding real unit norm initial vectors b and c exist, so that

$$\psi_{k,\mathbb{R},\mathbb{R}}(A) = \|p_b(A)b\| = \|p_c(A)c\|.$$

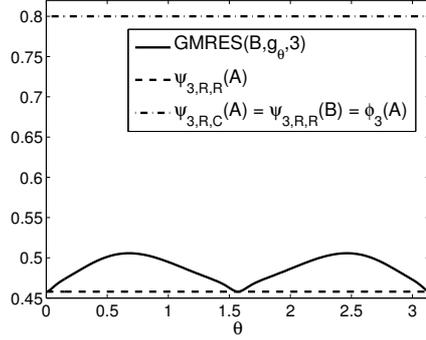


FIG. 6.1. Illustration of the value of (6.5) and different quantities from Theorem 6.3 for the Toh matrix $A(1.0, 0.1)$ in (4.1) and $k = 3$.

Note that since p_b and p_c are the uniquely determined GMRES polynomials that solve the problem (1.1) for the corresponding real initial vectors, it holds that

$$(6.4) \quad \|p_b(A)b\| < \|p(A)b\|, \quad \|p_c(A)c\| < \|p(A)c\|$$

for any polynomial $p \in \pi_{k,\mathbb{C}} \setminus \{p_b, p_c\}$.

Writing any complex vector $w \in \mathbb{C}^n$ in the form $w = (\cos \theta)u + \mathbf{i}(\sin \theta)v$, with $u, v \in \mathbb{R}^n$, $\|u\| = \|v\| = 1$, we get

$$\begin{aligned} \psi_{k,\mathbb{R},\mathbb{C}}^2(A) &= \max_{\substack{w \in \mathbb{C}^n \\ \|w\|=1}} \min_{p \in \pi_{k,\mathbb{R}}} \|p(A)b\|^2 \\ &= \max_{\substack{\theta \in \mathbb{R}, u, v \in \mathbb{R}^n \\ \|u\|=\|v\|=1}} \min_{p \in \pi_{k,\mathbb{R}}} (\cos^2 \theta \|p(A)u\|^2 + \sin^2 \theta \|p(A)v\|^2) \\ &\geq \max_{\theta \in \mathbb{R}} \min_{p \in \pi_{k,\mathbb{R}}} (\cos^2 \theta \|p(A)b\|^2 + \sin^2 \theta \|p(A)c\|^2) \\ &> (\cos^2 \theta) \psi_{k,\mathbb{R},\mathbb{R}}^2(A) + (\sin^2 \theta) \psi_{k,\mathbb{R},\mathbb{R}}^2(A) = \psi_{k,\mathbb{R},\mathbb{R}}^2(A), \end{aligned}$$

where the strict inequality follows from (6.4) and from the fact that $\|p(A)b\|^2$ and $\|p(A)c\|^2$ do not attain their minima for the same polynomial.

To demonstrate the strict inequality $\psi_{k,\mathbb{R},\mathbb{R}}(A) < \psi_{k,\mathbb{R},\mathbb{C}}(A)$ numerically we use the Toh matrix (4.1) with $\varepsilon = 0.1$ and $\omega = 1$, and $k = 3$. Let b and c be the corresponding two different worst-case initial vectors introduced in Section 4. We vary θ from 0 to π and compute the quantities

$$(6.5) \quad \min_{p \in \pi_{3,\mathbb{R}}} (\cos^2 \theta \|p(A)b\|^2 + \sin^2 \theta \|p(A)c\|^2) = \min_{p \in \pi_{3,\mathbb{R}}} \|p(B)g_\theta\|^2,$$

where

$$B = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \quad \text{and} \quad g_\theta = \begin{bmatrix} (\cos \theta)b \\ (\sin \theta)c \end{bmatrix}.$$

In Fig. 6.1 we can see clearly, that for $\theta \notin \{0, \pi/2, \pi\}$ the value of (6.5) is strictly larger than $\psi_{3,\mathbb{R},\mathbb{R}}(A) = 0.4579$ (dashed line). Numerical computations predict that $\psi_{3,\mathbb{R},\mathbb{R}}(A) = \psi_{3,\mathbb{C},\mathbb{C}}(A)$ for the Toh matrix. Finally, Lemma 6.2 and equation (6.3) imply $\psi_{3,\mathbb{R},\mathbb{C}}(A) = \psi_{3,\mathbb{R},\mathbb{R}}(B) = \varphi_3(A)$ (dash-dotted line).

Acknowledgements. We thank Michael Eiermann and two anonymous referees for very helpful corrections and suggestions.

REFERENCES

- [1] M. AFANASJEW, M. EIERMANN, O. G. ERNST, AND S. GÜTTEL, *A generalization of the steepest descent method for matrix functions*, Electron. Trans. Numer. Anal., 28 (2007/08), pp. 206–222.
- [2] V. FABER, W. JOUBERT, E. KNILL, AND T. MANTEUFFEL, *Minimal residual method stronger than polynomial preconditioning*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 707–729.
- [3] G. E. FORSYTHE, *On the asymptotic directions of the s -dimensional optimum gradient method*, Numer. Math., 11 (1968), pp. 57–76.
- [4] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, vol. 17 of Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 1997.
- [5] A. GREENBAUM AND L. GURVITS, *Max-min properties of matrix factor norms*, SIAM J. Sci. Comput., 15 (1994), pp. 348–358.
- [6] A. GREENBAUM AND L. N. TREFETHEN, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comput., 15 (1994), pp. 359–368.
- [7] W. JOUBERT, *On the convergence behavior of the restarted GMRES algorithm for solving nonsymmetric linear systems*, Numer. Linear Algebra Appl., 1 (1994), pp. 427–447.
- [8] ———, *A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems*, SIAM J. Sci. Comput., 15 (1994), pp. 427–439.
- [9] P. D. LAX, *Linear Algebra and its Applications*, Pure and Applied Mathematics (Hoboken), Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second ed., 2007.
- [10] J. LIESEN AND Z. STRAKOŠ, *Convergence of GMRES for tridiagonal Toeplitz matrices*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 233–251 (electronic).
- [11] J. LIESEN AND Z. STRAKOŠ, *Krylov Subspace Methods. Principles and Analysis*, Oxford University Press, Oxford, 2013.
- [12] J. LIESEN AND P. TICHÝ, *On best approximations of polynomials in matrices in the matrix 2-norm*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 853–863.
- [13] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, PA, second ed., 2003.
- [14] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
- [15] P. TICHÝ, J. LIESEN, AND V. FABER, *On worst-case GMRES, ideal GMRES, and the polynomial numerical hull of a Jordan block*, Electron. Trans. Numer. Anal., 26 (2007), pp. 453–473.
- [16] K.-C. TOH, *GMRES vs. ideal GMRES*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 30–36.
- [17] K.-C. TOH, M. J. TODD, AND R. H. TÜTÜNCÜ, *On the implementation and usage of SDPT3—a Matlab software package for semidefinite-quadratic-linear programming, version 4.0*, in Handbook on semidefinite, conic and polynomial optimization, vol. 166 of Internat. Ser. Oper. Res. Management Sci., Springer, New York, 2012, pp. 715–754.
- [18] K.-C. TOH AND L. N. TREFETHEN, *The Chebyshev polynomials of a matrix*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 400–419.
- [19] I. ZAVORIN, *Spectral factorization of the Krylov matrix and convergence of GMRES*, Tech. Report CS-TR-4309, Computer Science Department, University of Maryland, 2001.
- [20] I. ZAVORIN, D. P. O’LEARY, AND H. ELMAN, *Complete stagnation of GMRES*, Linear Algebra Appl., 367 (2003), pp. 165–183.