Electronic Transactions on Numerical Analysis.
Volume 20, pp. 180-197, 2005.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

ON THE WORST-CASE CONVERGENCE OF MR AND CG FOR SYMMETRIC POSITIVE DEFINITE TRIDIAGONAL TOEPLITZ MATRICES*

JÖRG LIESEN† AND PETR TICHÝ†

Abstract. We study the convergence of the minimal residual (MR) and the conjugate gradient (CG) method when applied to linear algebraic systems with symmetric positive definite tridiagonal Toeplitz matrices. Such systems arise, for example, from the discretization of one-dimensional reaction-diffusion equations with Dirichlet boundary conditions. Based on our previous results in [J. Liesen and P. Tichý, *BIT*, 44 (2004), pp. 79–98], we concentrate on the next-to-last iteration step, and determine the initial residuals and initial errors for the MR and CG method, respectively, that lead to the slowest possible convergence. By this we mean that the methods have made the least possible progress in the next-to-last iteration step. Using these worst-case initial vectors, we discuss which source term and boundary condition in the underlying reaction-diffusion equation are the worst in the sense that they lead to the worst-case initial vectors for the MR and CG methods. Moreover, we determine (or very tightly estimate) the worst-case convergence quantities in the next-to-last step, and compare these to the convergence quantities obtained from average (or unbiased) initial vectors. The spectral structure of the considered matrices allows us to apply our worst-case results for the next-to-last step to derive worst-case bounds also for other iteration steps. We present a comparison of the worst-case convergence quantities with the classical convergence bound based on the condition number of A, and finally we discuss the MR and CG convergence for the special case of the one-dimensional Poisson equation with Dirichlet boundary conditions.

Key words. Krylov subspace methods, conjugate gradient method, minimal residual method, convergence analysis, tridiagonal Toeplitz matrices, Poisson equation

AMS subject classifications. 15A09, 65F10, 65F20

1. Introduction.

This paper is concerned with the convergence analysis of Krylov subspace methods for solving linear algebraic systems of the form

$$Ax = b, \tag{1.1}$$

with a *symmetric positive definite* matrix $A \in \mathbb{R}^{n\times n}$, and a right hand side vector $b \in \mathbb{R}^n$. We obviously assume $n > 1$. Starting from an initial guess x_0, Krylov subspace methods compute the initial residual $r_0 = b - Ax_0$, and a sequence of approximate solutions (iterates) $x_1, x_2, \ldots$, such that the ith residual $r_i = b - Ax_i$ and the ith error $e_i = x - x_i$ are of the form

$$r_i = p_i(A) r_0 , \quad e_i = p_i(A) e_0 , \quad p_i \in \pi_i ,$$

where π_i denotes the set of polynomials of degree at most i and with value one at the origin. Two choices of conditions for determining the polynomials p_i have emerged as de facto standards.

In the minimal residual (MR) Krylov subspace method, the polynomial is chosen so that the Euclidean norm ($\|y\| = (y^T y)^{1/2}$) of the residuals is minimized,

$$\|r_i\| = \min_{p\in\pi_i} \|p(A) r_0\| \qquad \text{(MR)}. \tag{1.2}$$

There are several algorithms for implementing the MR method that try to exploit as much as possible from the properties of A. Examples are the conjugate residual (CR) method [18]

*Received January 15, 2005. Accepted for publication July 11, 2005. Recommended by D. Szyld.
†Institute of Mathematics, Technical University of Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany. (liesen@math.tu-berlin.de and tichy@math.tu-berlin.de). This research was supported by the Emmy Noether-Programm of the Deutsche Forschungsgemeinschaft. The work of the second author was also supported by the Grant Agency of Academy of Sciences of the Czech Republic under grant No. KJB1030306.

for symmetric positive definite A, the minimal residual (MINRES) method [17] symmetric and nonsingular A, and the generalized minimal residual (GMRES) method [19] for general nonsingular A.

In the orthogonal residual Krylov subspace method, the ith iterate x_i is determined such that the ith residual r_i is orthogonal to all previous residuals $r_0, \ldots, r_{i-1}$. A particular implementation for symmetric positive definite matrices A is the conjugate gradient (CG) method [8]. The symmetric positive definite matrix A defines a norm (A-norm, $\|y\|_A = (y^T A y)^{1/2}$) in which the errors are minimized,

$$\|e_i\|_A \;=\; \min_{p \in \pi_i} \|p(A)e_0\|_A \qquad \text{(CG)}. \tag{1.3}$$

The standard approach to analyze (1.2) and (1.3) is to exclude the influence of r_0 and e_0, and hence to consider the *worst-case convergence* instead of the convergence for the particular initial vectors. It is well known [4, 6, 9] that the (attainable) worst-case convergence quantities are given by

$$\max_{r_0 \neq 0} \min_{p \in \pi_i} \frac{\|p(A)r_0\|}{\|r_0\|} \;=\; \max_{e_0 \neq 0} \min_{p \in \pi_i} \frac{\|p(A)e_0\|_A}{\|e_0\|_A} \;=\; \min_{p \in \pi_i} \max_k |p(\lambda_k)|\,, \tag{1.4}$$

where λ_k, $k = 1, \ldots, n$, are the eigenvalues of A. The rightmost term in (1.4) depends in a nonlinear way on the eigenvalue distribution, and no explicit solution for this min-max approximation problem is known in general. Therefore, to analyze the worst-case convergence of the MR and CG methods one needs to *estimate* this min-max value. Such estimation can be based either on a suitable superset of the eigenvalues, or a suitable subset, where the first choice leads to an upper and the second to a lower bound on the worst-case convergence.

The standard choice of a superset of the discrete set of matrix eigenvalues is their convex hull $[\lambda_{\min}, \lambda_{\max}]$. Using scaled and shifted Chebyshev polynomials of the first kind on this interval, one can show the classical bound

$$\min_{p \in \pi_i} \max_k |p(\lambda_k)| \;\leq\; 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^i , \tag{1.5}$$

where $\kappa(A) = \lambda_{\max}/\lambda_{\min}$ is the condition number of A; see, e.g., [5, Theorem 3.1.1]. Because of (1.4), the term on the right hand side of (1.5) represents a bound on the relative residual norm $\|r_i\|/\|r_0\|$ for MR and the relative A-norm of the error $\|e_i\|_A/\|e_0\|_A$ for CG for each initial residual r_0 and each initial error e_0, respectively. The bound (1.5) is particularly useful in practical applications when only partial information about the spectrum of A is available or can be estimated. But one should be aware that this bound is obtained from a different kind of approximation problem than the one solved by the MR and CG methods (worst-case rather than for a specific r_0 or e_0, and continuous rather than discrete), and hence that it might provide misleading information about the actual convergence of these methods; see [12] for more details and references.

To obtain a lower bound on the worst-case convergence one can in principle consider any subset of the eigenvalues. As shown in [4, 13], for each subset of exactly $i+1$ distinct eigenvalues $\{\mu_1, \ldots, \mu_{i+1}\} \subseteq \{\lambda_1, \ldots, \lambda_n\}$,

$$\min_{p \in \pi_i} \max_k |p(\lambda_k)| \;\geq\; \min_{p \in \pi_i} \max_k |p(\mu_k)| \;=\; \left(\sum_{j=1}^{i+1} \prod_{\substack{l=1 \\ l \neq j}}^{i+1} \frac{|\mu_l|}{|\mu_l - \mu_j|} \right)^{-1} . \tag{1.6}$$

ETNA
Kent State University
etna@mcs.kent.edu

Apparently, for each choice of $i+1$ distinct eigenvalues $\mu_1, \ldots, \mu_{i+1}$, the right hand side of (1.6) represents an *explicit* lower bound on the worst-case convergence quantities. Moreover, in our case of real eigenvalues, there exists a subset of $i+1$ eigenvalues, for which the lower bound (1.6) is attained. Therefore, if the subset of $i+1$ eigenvalues is properly chosen, one can obtain a very good convergence estimate. Since this estimate of the worst-case convergence requires precise knowledge about at least some eigenvalues of A, its main use is in the analysis of model problems, where the eigenvalues are known explicitly.

In this paper we consider such a class of model problems, namely the linear systems with symmetric positive definite tridiagonal Toeplitz matrices A. Such systems arise, for example, in the discretization of one-dimensional reaction-diffusion equations. We focus on the *slowest possible convergence* of the MR and CG methods. By this we mean the situation when the worst-case convergence quantity is attained in the next-to-last iteration step. For this step the only possible subset $\{\mu_1, \ldots, \mu_{i+1}\}$ of the eigenvalues of A to be chosen in (1.6) is the set of all distinct eigenvalues of A, so that the solution of the min-max approximation problem is known explicitly. Based on our previous results in [13], we determine the worst possible initial data, i.e. the vectors r_0^w and e_0^w leading to the slowest possible convergence of the MR and CG method, respectively. Knowing the initial vector e_0^w explicitly, we identify source terms and boundary conditions in the one-dimensional reaction-diffusion equation that yield, after discretization, the slowest possible CG convergence. We also address the identification of such data for the MR method, which appears to be considerably more complicated than for CG. Moreover, we determine (or very tightly estimate) the worst-case convergence quantities in the next-to-last step, and compare these to the convergence quantities obtained from average (or unbiased) initial residuals as well as the classical convergence bound (1.5). The spectral structure of the considered matrices allows us to apply our worst-case results for the next-to-last step to derive worst-case bounds also for other iteration steps. Finally, we consider the case of one-dimensional Poisson equation, which is a popular model problem for the convergence analysis of Krylov subspace methods, in particular of CG; see, e.g., [1, 2, 15, 16].

We point out that the convergence of GMRES for nonsymmetric tridiagonal Toeplitz matrices is studied in [10]. The results in [10] hold explicitly for the highly nonnormal case, i.e. the case when a tridiagonal Toeplitz matrix can be considered a perturbed Jordan block. Hence the results presented in this paper are neither special cases nor generalizations of the results in [10].

The paper is organized as follows. Section 2 presents basic formulas for the next-to-last MR and CG iteration step. In Section 3 we focus on symmetric positive definite tridiagonal Toeplitz matrices that arise from the discretization of one-dimensional reaction-diffusion equations with Dirichlet boundary conditions, and study the MR and CG convergence quantities in the next-to-last step. Section 4 compares our results with known results for the Poisson equation model problem. Our conclusions are given in Section 5, and the Appendix lists all trigonometric formulas used in the proofs.

2. Formulas for the next-to-last MR and CG iteration step. Let a symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$ be given and denote by $A = Q\Lambda Q^T$ its eigendecomposition, where $Q^TQ = I$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. To avoid unnecessary technical complications we assume that *all eigenvalues of A are distinct*. Next, we parameterize the initial residual r_0 and the initial error e_0 by

$$r_0 = Q\,[\varrho_1, \ldots, \varrho_n]^T\,, \qquad e_0 = Q\,[\xi_1, \ldots, \xi_n]^T\,. \tag{2.1}$$

Note that, since $r_0 = Ae_0$, we have $\varrho_k = \lambda_k \xi_k$ for all $k = 1, \ldots, n$. Without loss of generality we restrict our analysis to vectors r_0 with $\varrho_k \neq 0$ for all $k = 1, \ldots, n$. In case

$d \geq 1$ coordinates ϱ_j are zero, the corresponding eigencomponents do not play any role, and hence the formulas for $i = n-1$ presented below will hold for $i = n-d-1$.

2.1. General results. As shown in [13, Theorem 2.1], the MR residual norm in the $(n-1)$st (next-to-last) iteration step is given by

$$\|r_{n-1}^{MR}\| = \left(\sum_{j=1}^{n} \left| \frac{L_j}{\varrho_j} \right|^2 \right)^{-1/2} = \left(\sum_{j=1}^{n} \left| \frac{L_j}{\lambda_j \xi_j} \right|^2 \right)^{-1/2}, \tag{2.2}$$

where

$$L_k \equiv \prod_{\substack{j=1 \\ j \neq k}}^{n} \frac{|\lambda_j|}{|\lambda_j - \lambda_k|}. \tag{2.3}$$

To obtain a similar result for the A-norm of the CG error, it suffices to realize that

$$\|p(A)e_0\|_A = \|p(A)A^{1/2}e_0\| \equiv \|p(A)\tilde{r}_0\|. \tag{2.4}$$

Hence the A-norm of the CG error can be seen as the MR residual norm, when MR is started with the initial residual $\tilde{r}_0 = A^{1/2}e_0$. Parameterizing $\tilde{r}_0$ by $\tilde{r}_0 = Q[\tilde{\varrho}_1, \ldots, \tilde{\varrho}_n]^T$, i.e. $\tilde{\varrho}_k = \lambda_k^{1/2}\xi_k = \lambda_k^{-1/2}\varrho_k$, we obtain

$$\|e_{n-1}^{CG}\|_A = \left(\sum_{j=1}^{n} \left| \frac{L_j}{\lambda_j^{1/2}\xi_j} \right|^2 \right)^{-1/2} = \left(\sum_{j=1}^{n} \left| \frac{\lambda_j^{1/2} L_j}{\varrho_j} \right|^2 \right)^{-1/2}. \tag{2.5}$$

The formulas (2.2) and (2.5) provide explicit a priori information about the next-to-last MR and CG convergence quantities in terms of the matrix eigenvalues and the coordinates of r_0 or e_0 in the matrix eigenvectors. To simplify the notation, we will write residuals and errors without superscript MR or CG. When we speak about residuals r_i, we always mean residuals r_i^{MR} of the MR method. Similarly, e_i always denotes the error e_i^{CG} of the CG method. The superscript can be now used to indicate the association of a residual or error with a particular initial residual or error.

2.2. Convergence quantities for different initial vectors. As described in the Introduction, we are interested in initial residuals and initial errors that lead to the maximal relative convergence quantities of the MR and CG method, respectively, in the next-to-last iteration step. We denote such a worst-case initial residual for the MR method by r_0^w, and the corresponding residual in the next-to-last step by r_{n-1}^w. In [13, Theorem 3.1] we show that

$$r_0^w = Q[\varrho_1^w, \ldots, \varrho_n^w]^T, \qquad |\varrho_k^w|^2 = \gamma L_k, \quad k = 1, \ldots, n, \tag{2.6}$$

where $\gamma > 0$ is any scaling factor, and that

$$\frac{\|r_{n-1}^w\|}{\|r_0^w\|} = \max_{r_0 \neq 0} \min_{p \in \pi_{n-1}} \frac{\|p(A)r_0\|}{\|r_0\|} = \left(\sum_{k=1}^{n} L_k \right)^{-1}. \tag{2.7}$$

Using the relation (2.4) and the definition of r_0^w it is not hard to see that the corresponding worst-case initial error e_0^w for CG is given by

$$e_0^w = Q[\xi_1^w, \ldots, \xi_n^w]^T, \qquad |\xi_k^w|^2 = \gamma \lambda_k^{-1} L_k \quad \text{for} \quad k = 1, \ldots, n, \tag{2.8}$$

ETNA
Kent State University
etna@mcs.kent.edu

where $\gamma > 0$ is any scaling factor, and that

$$\frac{\|e^w_{n-1}\|_A}{\|e^w_0\|_A} = \max_{e_0 \neq 0} \min_{p \in \pi_{n-1}} \frac{\|p(A)e_0\|_A}{\|e_0\|_A} = \left(\sum_{k=1}^{n} L_k \right)^{-1} . \tag{2.9}$$

We also consider the initial residual

$$r^u_0 = Q[\varrho^u_1, \dots, \varrho^u_n]^T , \qquad \varrho^u_k = 1, \quad k = 1, \dots, n. \tag{2.10}$$

The vector r^u_0 can be considered as a representative of the initial residuals which are uncorrelated with the matrix A, in the sense that their components in the eigenvectors of A are of (approximately) equal size. We call such vectors *unbiased* with respect to A. The MR method started with the initial residual (2.10) will produce, in the next-to-last iteration step, the residual vector r^u_{n-1}. Using (2.2), the relative MR residual norm is given by

$$\frac{\|r^u_{n-1}\|}{\|r^u_0\|} = \left(n \sum_{k=1}^{n} L_k^2 \right)^{-1/2} . \tag{2.11}$$

The CG method started with the initial residual r^u_0, i.e. with the initial error

$$e^u_0 = A^{-1} r^u_0 = Q[\xi^u_1, \dots, \xi^u_n]^T = Q[\lambda_1^{-1}, \dots, \lambda_n^{-1}]^T, \tag{2.12}$$

generates in the next-to-last iteration step the error e^u_{n-1}. Based on (2.5), the relative A-norm of this error is given by

$$\frac{\|e^u_{n-1}\|_A}{\|e^u_0\|_A} = \left(\sum_{k=1}^{n} \lambda_k L_k^2 \right)^{-1/2} \left(\sum_{k=1}^{n} \frac{1}{\lambda_k} \right)^{-1/2} . \tag{2.13}$$

The vector e^u_0 is by its definition correlated with the eigenvalue distribution of A and thus can be considered *biased*. We have deliberately made this choice to contrast the convergence quantities of MR and CG for the same initial residual.

3. Symmetric positive definite tridiagonal Toeplitz matrices. Consider the one-dimensional reaction-diffusion equation

$$-\,u''(z) + \sigma u(z) = f(z), \quad z \in (0, 1), \tag{3.1}$$

for some parameter $\sigma \geq 0$, with Dirichlet boundary conditions

$$u(0) = u_0, \quad u(1) = u_1 \,. \tag{3.2}$$

Then for each positive integer n, the central finite difference approximation of (3.1)–(3.2) on the uniform grid kh, $k = 1, \dots, n$, $h = (n+1)^{-1}$, leads to a linear system of the form

$$\underbrace{\begin{bmatrix} 2(1+\delta) & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2(1+\delta) \end{bmatrix}}_{A} x = \underbrace{h^2 \begin{bmatrix} f(h) \\ \vdots \\ \vdots \\ f(nh) \end{bmatrix} + \begin{bmatrix} u_0 \\ \\ \\ u_1 \end{bmatrix}}_{b} . \tag{3.3}$$

ETNA
Kent State University
etna@mcs.kent.edu

In the expression for A we have defined $\delta \equiv \sigma h^2/2$ for notational convenience.

The n distinct and positive eigenvalues λ_k, and the normalized eigenvectors q_k of A are given by

$$\lambda_k = 2(1+\delta) - 2\,\omega_k = 2\delta + 4\sin^2(k\pi h/2)\,, \quad \omega_k \equiv \cos(k\pi h)\,, \tag{3.4}$$

$$q_k = (2h)^{1/2}\,\left[\sin(k\pi h)\,, \sin(2k\pi h)\,, \ldots, \sin(nk\pi h)\right]^T\,, \quad k = 1, \ldots, n\,, \tag{3.5}$$

cf., e.g., [20, pp. 113–115]. We write the eigendecomposition of A as $A = Q\Lambda Q^T$, where $Q = [q_1, \ldots, q_n]$, and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

REMARK 3.1. We have chosen to derive our results for the tridiagonal Toeplitz matrix $A = \mathrm{tridiag}(-1, 2(1+\delta), -1)$ in (3.3) because of its direct relation to the differential equation (3.1)–(3.2). However, our results hold equally well for any symmetric tridiagonal Toeplitz matrix of the form $B = \mathrm{tridiag}(\beta, \alpha, \beta)$ with $\alpha = 2|\beta|(1+\delta) > 0$, for some $\delta > 0$. Obviously, $B = |\beta|\,\mathrm{tridiag}(\beta/|\beta|, 2(1+\delta), \beta/|\beta|)$. If $\beta < 0$, then $B = |\beta| A$, and if $\beta > 0$, then $B = |\beta| I^{\pm} A I^{\pm}$, where $I^{\pm} = \mathrm{diag}(1, -1, \ldots, (-1)^{n+1})$. In either case, A and B have the same set of orthogonal eigenvectors, and the eigenvalues B coincide with those of A up to a scaling by $|\beta|$. It is easy to check that all of our results are invariant under such scaling of the eigenvalues of A.

3.1. Connection with Chebyshev polynomials of the second kind. The relation of the eigenvalues of A given in (3.4) to the roots of the nth Chebyshev polynomial of the *second* kind, denoted by $U_n(z)$, will prove useful in our context. The polynomial $U_n(z)$ has degree n, and its n distinct roots are the values $\omega_k = \cos(k\pi h)$, $k = 1, \ldots, n$. Hence all roots are contained in the open interval $(-1, 1)$. The leading coefficient of $U_n(z)$ is 2^n, which means that $U_n(z)$ can be written as

$$U_n(z) = 2^n \prod_{k=1}^{n} (z - \omega_k)\;.$$

This relation shows that the product of all eigenvalues of A can be expressed as

$$\prod_{k=1}^{n} \lambda_k \;=\; 2^n \prod_{k=1}^{n} (1 + \delta - \omega_k) \;=\; U_n(1+\delta)\,. \tag{3.6}$$

Below we study how much the MR and CG convergence quantities change with changing δ. For this we first need to understand the behavior of $U_n(1+\delta)$ as a function of $\delta \geq 0$. To get a feeling of the growth of $U_n(z)$ outside the interval $(-1, 1)$, we use the alternative representation

$$U_n(z) = \frac{1}{2}\,\frac{(z + \sqrt{z^2-1})^{n+1} - (z - \sqrt{z^2-1})^{n+1}}{\sqrt{z^2-1}}\,, \tag{3.7}$$

see, e.g., [14, p. 15]. Using this formula, elementary real analysis shows that

$$U_n(1) = |U_n(-1)| = n + 1\,,$$

and that $U_n'(z) > 0$ for $z \geq 1$. In particular, $U_n(1+\delta)$ is positive and strictly increasing for $\delta \geq 0$. As shown by (3.7), $|U_n(z)|$ grows exponentially outside $(-1, 1)$. This is illustrated in Fig. 3.1, where we plot $U_n(z)/(n+1)$ for $n = 4, 6, 10$.

ETNA
Kent State University
etna@mcs.kent.edu

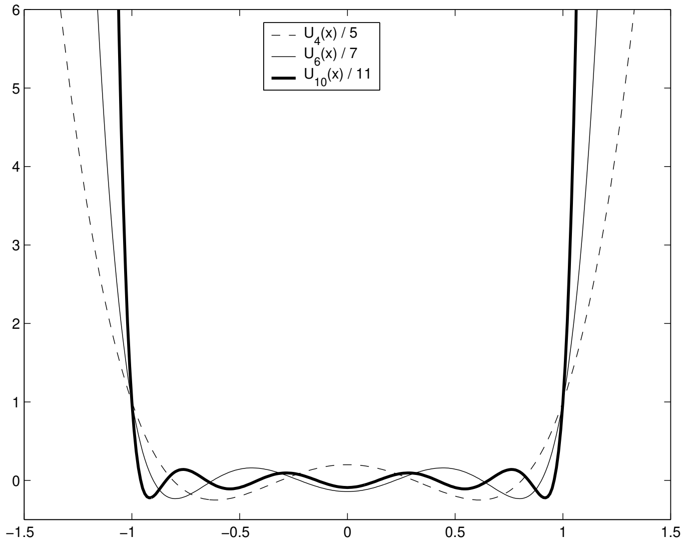

FIG. 3.1. $U_n(z)/(n+1)$ *for different* n.

3.2. Worst-case data. Our goal here is to characterize data (source term f and boundary conditions) in (3.1)–(3.2), that lead to the maximal relative convergence quantities in the next-to-last step when MR and CG with the initial guess $x_0 = 0$ are applied to the discretized system (3.3). Our main tools are the parameterizations (2.6) and (2.8) of the worst-case initial vectors r_0^w and e_0^w, which we evaluate explicitly using the known eigendecomposition of A, and then translate back into data for (3.1)–(3.2). The vectors r_0^w and e_0^w depend on the terms L_k, which are characterized by the following lemma.

LEMMA 3.2. *Suppose that* $\lambda_1, \ldots, \lambda_n$ *are given by* (3.4) *for some* $\delta \geq 0$. *Then* L_k *as defined in* (2.3) *satisfies*

$$L_k = h\, U_n\,(1+\delta)\; \frac{\sin^2(k\pi h)}{\delta + 2\sin^2\left(\frac{k\pi h}{2}\right)}\,. \tag{3.8}$$

In particular, for $\delta = 0$,

$$L_k \;=\; 2\cos^2\left(\frac{k\pi h}{2}\right). \tag{3.9}$$

Proof. The denominator of L_k can be written as

$$\begin{aligned}\prod_{\substack{j=1\\ j\neq k}}^{n} |\lambda_j - \lambda_k| \;&=\; \prod_{\substack{j=1\\ j\neq k}}^{n} |2\,\omega_k - 2\,\omega_j| \;=\; 2^{2n-2}\prod_{\substack{j=1\\ j\neq k}}^{n}\left|\sin^2\left(\frac{jh\pi}{2}\right) - \sin^2\left(\frac{kh\pi}{2}\right)\right| \\ &=\; \frac{n+1}{2\sin^2(k\pi h)}\,, \end{aligned}\tag{3.10}$$

cf. identity (A.1). According to (2.3), (3.6) and (3.10),

$$L_k \;=\; \frac{U_n\,(1+\delta)}{\lambda_k}\cdot\frac{2\sin^2(k\pi h)}{n+1} \;=\; h\,U_n\,(1+\delta)\;\frac{\sin^2(k\pi h)}{\delta + 2\sin^2\left(\frac{k\pi h}{2}\right)}\,. \tag{3.11}$$

ETNA
Kent State University
etna@mcs.kent.edu

The relation (3.9) for $\delta = 0$ follows immediately from $U_n(1) = n+1 = h^{-1}$ and $\sin(k\pi h) = 2\sin(k\pi h/2)\cos(k\pi h/2)$. $\square$

Now consider the parameterization of e_0^w given in (2.8). Clearly, for any $\gamma > 0$, the set of coefficients

$$\xi_k^w \;\equiv\; \left(\gamma \lambda_k^{-1} L_k\right)^{1/2}, \quad k = 1, \ldots, n, \tag{3.12}$$

leads to a worst-case initial error $e_0^w = Q[\xi_1^w, \ldots, \xi_n^w]^T$ for CG. If CG is started with initial guess $x_0 = 0$, then e_0^w represents the solution, and Ae_0^w the right hand side of a linear system that leads to the maximal relative A-norm of the error in the next-to-last iteration step.

Using the coefficients (3.12), and the explicit form of L_k in (3.11),

$$\begin{aligned}\lambda_k \xi_k^w &= \lambda_k \left(\gamma \lambda_k^{-1} \frac{U_n\,(1+\delta)}{\lambda_k} \cdot \frac{2\sin^2(k\pi h)}{n+1}\right)^{1/2} \\ &= \left(\gamma\, 2h\, U_n\,(1+\delta)\right)^{1/2} \sin(k\pi h)\,,\end{aligned}$$

and, therefore,

$$\begin{aligned}Ae_0^w &= (Q\Lambda Q^T)\,(Q[\xi_1^w, \ldots, \xi_n^w]^T) \\ &= Q\,[\lambda_1\xi_1^w, \ldots, \lambda_n\xi_n^w]^T \\ &= \left(\gamma\, 2h\, U_n\,(1+\delta)\right)^{1/2} Q\,[\sin(\pi h), \ldots, \sin(n\pi h)]^T \\ &= \left(\gamma\, 2h\, U_n\,(1+\delta)\right)^{1/2} Q\,(2h)^{-1/2}\, q_1 \\ &= \left(\gamma U_n\,(1+\delta)\right)^{1/2} [1, 0, \ldots, 0]^T. \end{aligned} \tag{3.13}$$

Since $\gamma > 0$ can be chosen arbitrarily, we conclude that any right hand side vector b that is a positive multiple of the first unit vector leads to the worst possible relative A-norm of the error in the next-to-last step of CG (with $x_0 = 0$) for the linear system $Ax = b$ given by (3.3). The convergence of CG (with $x_0 = 0$) for $Ax = b$ is obviously the same as for $Ax = -b$, and therefore any negative multiple of the first unit vector is a worst-case right hand side in the just described sense as well.

Instead of the coefficients (3.12) we may define

$$\xi_k^w \;\equiv\; (-1)^{k+1}\left(\gamma \lambda_k^{-1} L_k\right)^{1/2}, \quad k = 1, \ldots, n. \tag{3.14}$$

Then, using $(-1)^{k+1}\sin(k\pi h) = \sin(nk\pi h)$, we obtain

$$\lambda_k \xi_k^w \;=\; \left(\gamma\, 2h\, U_n\,(1+\delta)\right)^{1/2} \sin(nk\pi h)\,.$$

A computation analogous to the one leading to (3.13) shows that, for the initial error e_0^w defined by the coefficients (3.14),

$$Ae_0^w \;=\; \left(\gamma U_n\,(1+\delta)\right)^{1/2} [0, \ldots, 0, 1]^T, \tag{3.15}$$

i.e., any nonzero multiple of the nth unit vector also is a worst-case right hand side for CG.

Both examples show that the right hand sides leading to the very unfavorable convergence behavior of CG may look rather unsuspicious at first sight. In terms of the differential equation (3.1)–(3.2), the worst possible relative A-norm of the next-to-last error in CG (for $x_0 = 0$) is obtained simply by

$$f = 0 \quad \text{and} \quad u_0 = c,\; u_1 = 0, \quad \text{or} \quad u_0 = 0,\; u_1 = c\,, \tag{3.16}$$

ETNA
Kent State University
etna@mcs.kent.edu

for any nonzero constant c.

As shown in (2.4), CG for the initial error e_0^w defined by (3.12) is equivalent to MR for the initial residual $A^{1/2}e_0^w$ that can be written in the form

$$\begin{aligned} A^{1/2}e_0^w &= A^{-1/2}Ae_0^w \\ &= (\gamma U_n\,(1+\delta))^{1/2}\; A^{-1/2}[1,0,\ldots,0]^T \\ &= (\gamma U_n\,(1+\delta))^{1/2}\; Q\Lambda^{-1/2}q_1\,. \end{aligned}$$

Therefore, any nonzero multiple of the vector $r_0^w \equiv Q\Lambda^{-1/2}q_1$ leads to the worst-case relative residual norm in the next-to-last MR step. Obviously, the coordinates of r_0^w in the eigenvectors of A are given by

$$\varrho_k^w = [2\delta + 4\sin^2(k\pi h/2)]^{-1/2}\sin(k\pi h), \quad k = 1,\ldots,n\,. \tag{3.17}$$

Because of the complicated form of the ϱ_k^w, no simple expression for the vector $r_0^w = Q[\varrho_1^w,\ldots,\varrho_n^w]^T$ exists in general. An exception for which r_0^w can be found in a relatively simple form is the case $\delta = 0$, where $\varrho_k^w = \cos(k\pi h/2)$, and the jth entry of r_0^w, denoted by $r_{0,j}^w$ for $j = 1,\ldots,n$, satisfies

$$r_{0,j}^w \;=\; (2h)^{1/2}\frac{\sin(j\pi h)}{\cos\left(\frac{\pi h}{2}\right) - \cos(j\pi h)}\,. \tag{3.18}$$

As (3.18) indicates, for MR it is not as straightforward as for CG to find data for (3.1)–(3.2) that leads to the worst case in the next-to-last step. For more details and a proof of (3.18) we refer to [11].

3.3. Worst-case and unbiased convergence quantities. After having characterized the worst-case initial vectors r_0^w and e_0^w for the system (3.3), we next evaluate the corresponding convergence quantity (2.7) and compare it to the quantities (2.11) and (2.13) resulting from the initial vectors r_0^u and e_0^u. We start with deriving bounds on (2.7) and (2.11).

THEOREM 3.3. *Suppose that MR is applied to a system of the form (3.3), and the initial residual is either r_0^w or r_0^u. Then*

$$3^{-1}\,\frac{2+\delta}{U_n\,(1+\delta)} \;<\; \frac{\|r_{n-1}^u\|}{\|r_0^u\|} \;<\; \frac{\|r_{n-1}^w\|}{\|r_0^w\|} \;\le\; 3\,\frac{2+\delta}{U_n\,(1+\delta)}\,. \tag{3.19}$$

In particular, for $\delta = 0$,

$$\frac{1}{n}\sqrt{\frac{2}{3}} \;<\; \sqrt{\frac{2}{3n^2-n}} \;=\; \frac{\|r_{n-1}^u\|}{\|r_0^u\|} \;<\; \frac{\|r_{n-1}^w\|}{\|r_0^w\|} \;=\; \frac{1}{n}\,. \tag{3.20}$$

Proof. We first prove (3.19). The middle inequality is trivial. To show the leftmost inequality it suffices to use the relation (2.11) and to find an upper bound on the sum of the L_k^2. Using (3.8) and (A.4),

$$\begin{aligned} \sum_{k=1}^n L_k^2 \;&\le\; \frac{U_n^2\,(1+\delta)}{(n+1)^2(\frac{\delta}{2}+1)^2}\sum_{k=1}^n\frac{\sin^4(k\pi h)}{4\sin^4\left(\frac{k\pi h}{2}\right)} \\ &=\; \frac{16\,U_n^2\,(1+\delta)}{(n+1)^2(\delta+2)^2}\sum_{k=1}^n\cos^4\left(\frac{k\pi h}{2}\right) \\ &=\; \frac{(6\,n-2)\,U_n^2\,(1+\delta)}{(n+1)^2(\delta+2)^2}\,. \end{aligned} \tag{3.21}$$

Then (2.11) implies

$$\left(n \sum_{k=1}^{n} L_k^2\right)^{-1/2} \geq \frac{(n+1)(\delta+2)}{\sqrt{(6\,n-2)n}\, U_n\,(1+\delta)} > \frac{1}{3}\,\frac{\delta+2}{U_n\,(1+\delta)}\,.$$

Next note that, using (A.3),

$$\begin{aligned}\sum_{k=1}^{n} L_k &\geq \frac{U_n\,(1+\delta)}{\delta+2} \sum_{k=1}^{n} \frac{\sin^2(k\pi h)}{n+1} = \frac{1}{2}\,\frac{U_n\,(\delta+1)}{\delta+2}\,\frac{n}{n+1}\\ &\geq \frac{1}{3}\,\frac{U_n\,(\delta+1)}{\delta+2}\,,\end{aligned} \tag{3.22}$$

and thus the rightmost inequality in (3.19) follows from applying (3.22) to (2.7).

For $\delta = 0$ we have

$$\sum_{k=1}^{n} L_k = 2 \sum_{k=1}^{n} \cos^2\left(\frac{k\pi h}{2}\right) = n\,, \tag{3.23}$$

cf. (A.3), and

$$\begin{aligned}\sum_{k=1}^{n} L_k^2 &= \frac{U_n^2\,(1)}{(n+1)^2} \sum_{k=1}^{n} \frac{\sin^4(k\pi h)}{4\sin^4\left(\frac{k\pi h}{2}\right)}\\ &= 4 \sum_{k=1}^{n} \cos^4\left(\frac{k\pi h}{2}\right) = \frac{3\,n-1}{2}\,,\end{aligned} \tag{3.24}$$

cf. (A.4). Substituting (3.23) and (3.24) into (2.7) and (2.11), we obtain (3.20). ☐

Since $\|r_{n-1}^w\|/\|r_0^w\| = \|e_{n-1}^w\|_A/\|e_0^w\|_A$ (compare (2.7) and (2.9)) the theorem also characterizes $\|e_{n-1}^w\|_A/\|e_0^w\|_A$, the next-to-last worst-case relative A-norm of the error for CG.

The rightmost equation in (3.20) shows that, for $\delta = 0$, MR in the worst case decreases the relative residual norm in the first $n-1$ iteration steps only to n^{-1}. On the other hand, since $\|r_{n-1}^w\|/\|r_0^w\| \approx (1+\delta)/U_n(1+\delta)$ for all δ, the next-to-last worst-case MR residual norm decreases exponentially with increasing δ, and hence increasing diagonal dominance of A. Moreover, Theorem 3.3 shows that the progress MR has made in the next-to-last iteration step for the unbiased initial residual r_0^u is at most a *constant factor* (less than $1/9$) apart from the worst case. In general the two cases may differ by a factor of up to $n^{1/2}$; see [13, Section 5], [7, Section 5].

The spectral structure of A allows to use the worst-case convergence result for the next-to-last step in Theorem 3.3 to obtain a worst-case convergence bound also for other iteration steps.

COROLLARY 3.4. *Suppose that the positive integer m divides $n+1$. Then for all $i \equiv (n+1)/m - 2 > 1$,*

$$\max_{r_0 \neq 0} \min_{p \in \pi_i} \frac{\|p(A)r_0\|}{\|r_0\|} = \max_{e_0 \neq 0} \min_{p \in \pi_i} \frac{\|p(A)e_0\|_A}{\|e_0\|_A} > 3^{-1}\,\frac{2+\delta}{U_{i+1}\,(1+\delta)}\,. \tag{3.25}$$

ETNA
Kent State University
etna@mcs.kent.edu

Proof. Consider the subset $\{\mu_1, \ldots, \mu_{i+1}\} \subseteq \{\lambda_1, \ldots, \lambda_n\}$ of $i+1$ eigenvalues of A given by

$$\mu_j = 2\left(1 + \delta - \cos\left(\frac{j\,\pi}{i+2}\right)\right), \qquad j = 1, \ldots, i+1\,.$$

It is easy to see that the set $\{\mu_1, \ldots, \mu_{i+1}\}$ consists of the $i+1$ distinct eigenvalues of $A_{i+1} \equiv \mathrm{tridiag}(-1, 2(1+\delta), -1) \in \mathbb{R}^{(i+1)\times(i+1)}$. Then

$$\begin{aligned}
\max_{r_0 \neq 0} \min_{p \in \pi_i} \frac{\|p(A) r_0\|}{\|r_0\|} &= \min_{p \in \pi_i} \max_k |p(\lambda_k)| \\
&\geq \min_{p \in \pi_i} \max_k |p(\mu_k)| \\
&= \max_{r_0 \neq 0} \min_{p \in \pi_i} \frac{\|p(A_{i+1}) r_0\|}{\|r_0\|} \\
&> 3^{-1} \frac{2+\delta}{U_{i+1}\,(1+\delta)}\,,
\end{aligned}$$

where the final lower bound results from applying Theorem 3.3 to the linear system with A_{i+1}. ∎

For example, in case $n = 99$, the lower bound (3.25) would apply in the steps $i = 2, 8, 18, 23, 48, 98$. Hence in addition to just the lower bound on $\|r^w_{n-1}\|/\|r^w_0\|$ in (3.19), which corresponds to (3.25) for $i = n-1$, we get additional lower bounds particularly for the earlier phase of the iteration.

Theorem 3.3 does not characterize (2.13), i.e. the case of CG for the initial error e^u_0. This is done in the following result.

THEOREM 3.5. *Suppose that CG is applied to a system of the form (3.3), and the initial error is e^u_0. Then*

$$3^{-1} \frac{\delta}{U_n\,(1+\delta)} < \frac{\|e^u_{n-1}\|_A}{\|e^u_0\|_A} < 3\,\frac{2+\delta}{U_n\,(1+\delta)}\,. \tag{3.26}$$

For $\delta < 1/4$,

$$3^{-1} \frac{\delta + 2}{n^{1/2} U_n\,(1+\delta)} < \frac{\|e^u_{n-1}\|_A}{\|e^u_0\|_A}\,, \tag{3.27}$$

and for $\delta = 0$,

$$\frac{\|e^u_{n-1}\|_A}{\|e^u_0\|_A} = \frac{\sqrt{6}}{\sqrt{n(n+1)(n+2)}} > n^{-3/2}\,. \tag{3.28}$$

Proof. The second inequality in (3.26) follows easily from (3.19). We prove the first inequality. Using Cauchy's inequality we obtain, cf. (2.13),

$$\frac{\|e^u_0\|_A^2}{\|e^u_{n-1}\|_A^2} \leq \left(\sum_{k=1}^n L_k^4\right)^{1/2} \left(\sum_{k=1}^n \lambda_k^2\right)^{1/2} \left(\sum_{k=1}^n \frac{1}{\lambda_k}\right)\,. \tag{3.29}$$

ETNA
Kent State University
etna@mcs.kent.edu

Since λ_n is the largest eigenvalue,

$$\begin{aligned} \left(\sum_{k=1}^{n} \lambda_k^2\right)^{1/2} \left(\sum_{k=1}^{n} \frac{1}{\lambda_k}\right) &< n^{1/2}\lambda_n \left(\sum_{k=1}^{n} \frac{1}{\lambda_k}\right) \;<\; n^{3/2}\frac{\lambda_n}{\lambda_1} \\ &= n^{3/2}\frac{1+\delta+\omega_1}{1+\delta-\omega_1} \\ &< n^{3/2}\frac{2+\delta}{\delta}\,. \end{aligned} \tag{3.30}$$

It remains to find a bound on the sum of the L_k^4. Using (3.8) and (A.5),

$$\begin{aligned} \sum_{k=1}^{n} L_k^4 &\le \frac{U_n^4(1+\delta)}{(n+1)^4(\frac{\delta}{2}+1)^4} \sum_{k=1}^{n} \frac{\sin^8(k\pi h)}{2^4\sin^8\left(\frac{k\pi h}{2}\right)} \\ &= 2^8 \frac{U_n^4(1+\delta)}{(n+1)^4(\delta+2)^4} \sum_{k=1}^{n} \cos^8\left(\frac{k\pi h}{2}\right) \\ &< 3^4 \frac{n\,U_n^4(1+\delta)}{(n+1)^4(\delta+2)^4}\,. \end{aligned} \tag{3.31}$$

From (3.29)–(3.31) we now obtain (3.26).

Now consider the case $\delta < 1/4$. Then

$$\begin{aligned} \left(\sum_{k=1}^{n} \lambda_k^2\right)^{1/2} \sum_{k=1}^{n} \frac{1}{\lambda_k} &= \left(\sum_{k=1}^{n}(1+\delta-\omega_k)^2\right)^{1/2} \sum_{k=1}^{n} \frac{1}{1+\delta-\omega_k} \\ &< \left(\sum_{k=1}^{n}(5/4-\omega_k)^2\right)^{1/2} \sum_{k=1}^{n} \frac{1}{1-\omega_k} \\ &= \left(\frac{33}{16}n-\frac{1}{2}\right)^{1/2} \sum_{k=1}^{n} \frac{1}{2\sin^2\left(\frac{k\pi h}{2}\right)} \\ &< \left(\frac{36}{16}n\right)^{1/2} \frac{n(n+2)}{3} \\ &= \frac{n^{1/2}n(n+2)}{2} \\ &< \frac{n^{1/2}(n+1)^2}{2}\,, \end{aligned} \tag{3.32}$$

where we have used the identities (A.7) and (A.8). Then (3.27) follows from (3.29), (3.31) and (3.32).

For $\delta = 0$,

$$\begin{aligned} \frac{\|e_0^u\|_A^2}{\|e_{n-1}^u\|_A^2} &= \left(\sum_{k=1}^{n} 4\sin^2\left(\frac{k\pi h}{2}\right) 4\cos^4\left(\frac{k\pi h}{2}\right)\right)\left(\sum_{k=1}^{n} \frac{1}{4\sin^2\left(\frac{k\pi h}{2}\right)}\right) \\ &= (n+1)\left(\frac{n(n+2)}{6}\right)\,, \end{aligned}$$

where we have used (A.6) and (A.7). $\square$

A comparison of Theorems 3.3 and 3.5 shows that, for small δ,

$$\text{(MR)} \qquad \frac{\|r_{n-1}^u\|}{\|r_0^u\|} \;\approx\; n^{1/2}\,\frac{\|e_{n-1}^u\|_A}{\|e_0^u\|_A} \qquad \text{(CG)}\,.$$

For larger δ, this difference is much less pronounced, and these MR and CG quantities are at most a small constant apart from each other.

3.4. Comparison of the worst-case bound and the classical bound. We next compare our worst-case convergence results in Theorem 3.3 with the classical convergence bound (1.5),

$$\min_{p\in\pi_i}\max_k |p(\lambda_k)| \;\le\; 2\nu^i\,, \quad i=0,\ldots,n-1\,, \tag{3.33}$$

where $\nu \equiv (\sqrt{\kappa(A)}-1)/(\sqrt{\kappa(A)}+1) < 1$, for $i=n-1$.

For our comparison we express $U_n(1+\delta)$ in terms of the condition number of A, which is given by $\kappa(A)=\lambda_n/\lambda_1$. First note that, by (3.4),

$$1+\delta \;=\; \omega_1\,\frac{\lambda_n+\lambda_1}{\lambda_n-\lambda_1} \;=\; \omega_1\,\frac{\kappa(A)+1}{\kappa(A)-1} \;\equiv\; \omega_1\,\tau\,.$$

Next,

$$\tau-\sqrt{\tau^2-1} \;=\; \frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1} \;\equiv\; \nu\,, \qquad \tau+\sqrt{\tau^2-1} \;=\; \nu^{-1}\,, \tag{3.34}$$

which, inserted into (3.7), yields

$$U_n(\tau) \;=\; \frac{\nu^{n+1}-\nu^{-(n+1)}}{\nu-\nu^{-1}}\,. \tag{3.35}$$

Since $U_n(z)$ is strictly monotonically increasing for $z\ge 1$, and $\omega_1 \lesssim 1$,

$$U_n(1+\delta) \;\lesssim\; U_n(\tau) \;=\; \nu^{-n}+\nu^{-n+2}+\nu^{-n+4}+\ldots+\nu^n\,, \tag{3.36}$$

where "$\lesssim$" means that the inequality is close. In the notation established above,

$$2\nu^{n-1} \;\ge\; \frac{\|r_{n-1}^w\|}{\|r_0^w\|} \;=\; \frac{\|e_{n-1}^w\|_A}{\|e_0^w\|_A} \tag{3.37}$$

$$\gtrsim\; \frac{\|r_{n-1}^u\|}{\|r_0^u\|} \;\approx\; \frac{4}{\omega_1}\,\frac{2+\delta}{U_n\,(1+\delta)} \tag{3.38}$$

$$\gtrsim\; \frac{4\tau}{U_n(\tau)} \tag{3.39}$$

$$\gtrsim\; \frac{2}{\nu\,U_n(\tau)} \;=\; \frac{2\,\nu^{n-1}}{1+\nu^2+\ldots+\nu^{2(n-1)}+\nu^{2n}}\,. \tag{3.40}$$

In (3.37) we use (3.33) for $i=n-1$, and in (3.38) we use (3.19), where the unimportant multiplicative factor (between $1/3$ and 3) was replaced by $4/\omega_1$ for convenience. Next, in (3.39) we use (3.36) as well as the relation $\tau=(1+\delta)/\omega_1$, from which we receive (3.40) using (3.36) and the inequality $2\,\tau \ge \nu^{-1} \ge \tau$.

ETNA
Kent State University
etna@mcs.kent.edu

The main point in this derivation is that the actual convergence quantities on the right hand side of the inequality in (3.37) are always quite close to (3.40), i.e.

$$\frac{\|r^w_{n-1}\|}{\|r^w_0\|} = \frac{\|e^w_{n-1}\|_A}{\|e^w_0\|_A} \approx \frac{2\,\nu^{n-1}}{1+\nu^2+\ldots+\nu^{2(n-1)}+\nu^{2n}}\,.$$

The tightness of the *upper* bound (3.37) to the actual convergence quantities therefore depends on the size of ν, and hence on $\kappa(A)$, which for a fixed matrix size n is a strictly decreasing function of the parameter $\delta \geq 0$.

For small $\kappa(A)$ (or δ bounded away from zero), the difference between (3.37) and (3.40) is small, i.e. the classical bound provides accurate information about the actual convergence quantities of CG and MR in (3.37) and (3.38). On the other hand, when $\kappa(A)$ is large (or δ is close to zero), then the lower bound (3.40), and with it the CG and MR convergence quantities will be smaller (up to the factor n^{-1}) than predicted by the classical upper bound (3.37). In the limiting case $\delta = 0$,

$$\min_{p\in\pi_{n-1}} \max_{1\leq k\leq n} |p(\lambda_k)| = \frac{1}{n} \ll 2\nu^{n-1} \overset{n\to\infty}{\longrightarrow} 2e^{-\pi}\,.$$

This clearly demonstrates that, for reasonably large n, the classical bound (3.33) cannot describe the worst-case convergence values of CG or MR in later iterations. Asymptotically (for $n \to \infty$) the weakness of the classical bound in this context has also been noticed before by Axelsson [1, Example 13.7] and others.

4. Poisson equation. Now we consider the case of one-dimensional Poisson equation with Dirichlet boundary conditions, i.e. the problem (3.1)–(3.2) with $\sigma = 0$. Then $\delta = 0$ and the corresponding system matrix in (3.3) is $A = \mathrm{tridiag}(-1, 2, -1)$. In this case, simple explicit expressions for r^w_0 as well as e^w_0 are known (see Section 3.2). Moreover, we have determined the exact MR and CG convergence quantities in the next-to-last step for the worst-case as well as the unbiased initial vectors (see Theorems 3.3 and 3.5). In addition, it is possible, in this particular case and for special starting vectors including the ones considered in this paper, to determine the whole MR and CG convergence curve a priori. In the following we recall known results from [15] for the unbiased case, and state (without proof) a new convergence result for the worst case.

Assuming that $x_0 = 0$, and hence $e_0 = x$, the papers [15, 16] present exact analytic expressions for the relative A-norm of the CG errors for solutions of the form

$$x^{(s)} = Q[\xi^{(s)}_1, \ldots, \xi^{(s)}_n]^T\,, \qquad \xi^{(s)}_k = \sin^{-s}\left(\frac{k\pi h}{2}\right)\,, \tag{4.1}$$

for some parameter $s \in \mathbb{N}_0$. Two of these solutions are of particular interest in our context. A simple calculation shows that $x^{(2)} = 4e^u_0$ as defined in (2.12). Moreover, $A^{1/2}x^{(1)} = 2r^u_0$, where r^u_0 is defined in (2.10). Using these relations and the exact analytic convergence curves derived in [15] gives the following result.

PROPOSITION 4.1. *Suppose that CG and MR are applied to the system (3.3) with $\delta = 0$, and the respective initial error and residual are given by e^u_0 and r^u_0. Then the resulting CG errors e^u_i and MR residuals r^u_i, $i = 0, \ldots, n$, satisfy*

$$\frac{\|e^u_i\|_A}{\|e^u_0\|_A} = \left[\frac{(n-i)^3+3(n-i)^2+2(n-i)}{n(n+1)(n+2)}\right]^{1/2} \equiv \varphi_C(i)\,, \tag{4.2}$$

$$\frac{\|r^u_i\|}{\|r^u_0\|} = \left[\frac{(n-i)+(n-i)^2}{n(n+1)+2ni(n-i)}\right]^{1/2} \equiv \varphi_M(i)\,. \tag{4.3}$$

An elementary computation using (4.2) shows that

$$\frac{\varphi_C(i)}{\varphi_C(i-1)} = \left(\frac{n-i}{n-i+3}\right)^{1/2}, \qquad i=1,\ldots,n,$$

which represents a strictly decreasing function of the iteration step i. The "superlinear" behavior of $\varphi_C(i)$ can be related to the distribution of the eigenvector coordinates of the initial error e_0^u. As proved asymptotically by Beckermann and Kuijlaars [2], CG may for the model problem (3.3) with $\delta = 0$ converge superlinearly, when the initial error exhibits a certain distribution of eigencomponents that is far from an equilibrium distribution. This appears to be the case in our example, where e_0^u is *biased*, cf. (2.12).

Using the same techniques as in [15] based on Lagrange multipliers, it is also possible to determine the exact values of the relative A-norm of the error in every step of CG with the initial error e_0^w. This technique is quite involved, and the full proof would take us several pages to state. The final result is the following,

$$\frac{\|e_i^w\|_A}{\|e_0^w\|_A} = \left[\frac{n-i}{n\,(i+1)}\right]^{1/2} \equiv \varphi_W(i), \quad i=0,\ldots,n. \tag{4.4}$$

Because of the equivalence (2.4) between CG and MR, the relative MR residual norms for the initial residual r_0^w also satisfy $\|r_i^w\|/\|r_0^w\| = \varphi_W(i)$. Note that

$$\varphi_M(i) < \varphi_W(i) < \sqrt{2}\,\varphi_M(i), \quad i=1,\ldots,n-1. \tag{4.5}$$

Obviously, the worst-case convergence value (1.4) of CG and MR at each step i must be larger than (or equal) to any other attainable convergence value. Hence the maximum of the three convergence curves $\varphi_C(i)$, $\varphi_M(i)$ and $\varphi_W(i)$ forms a lower bound on the worst-case value,

$$\min_{p\in\pi_i}\max_k |p(\lambda_k)| \geq \max\{\varphi_C(i), \varphi_M(i), \varphi_W(i)\}, \quad i=0,\ldots,n-1. \tag{4.6}$$

Figure 4.1 illustrates the above results for the model problem (3.3) with $n = 120$ and $\delta = 0$. The computations were performed in MATLAB [21], on an AMD Athlon XP 2100+ personal computer with machine precision $\varepsilon \sim 10^{-16}$.

As predicted by (4.5), the curves $\varphi_M(i)$ (dashed dotted) and $\varphi_W(i)$ (solid) are very close. The left hand side of (4.6) (bold) was computed by the function `cheby0` of the semidefinite programming package SDPT3 [22]. Except for the last few steps, the maximum on the right hand side of (4.6) is given by $\varphi_C(i)$ (dashed). Overall, the bound (4.6) is quite tight. The bound (3.33) is tight in step i, if there exist $i-1$ eigenvalues of A, that closely approximate extrema of the ith scaled and shifted Chebyshev polynomial of the first kind. In our example this is not the case for the later phase of the iteration, where the two sides of (3.33) differ significantly.

As mentioned above, MR with the right hand side r_0^w (we used $x_0 = 0$ for MR and CG) and CG with the right hand side Ae_0^w have the same convergence curve given by $\varphi_W(i)$ (solid). However, the curves of MR with the right hand side Ae_0^w (dotted) and CG with the right hand side r_0^w (dashed dotted; coincides with $\varphi_M(i)$) differ by orders of magnitude from each other. Hence a right hand side that leads to the worst-case convergence for one method does not lead (in general) to similar convergence for the other method.

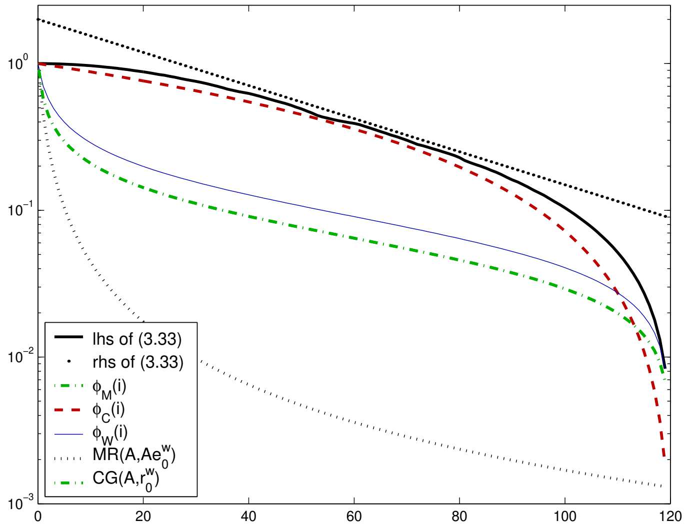

FIG. 4.1. *CG and MR convergence curves, and both sides of* (3.33).

5. Conclusions. In this paper we have applied our previous results in [13] to study the convergence of the CG and MR methods for linear systems with symmetric positive definite tridiagonal Toeplitz matrices. The structure of the matrix spectra allowed us to answer the questions how slow the convergence of the iterative solvers might possibly be for the considered model problems, which initial vectors lead to the maximal convergence quantity in the next-to-last iteration step, and how much the convergence quantity in this case differs from an “average” (or unbiased) case. We also were able to derive lower bounds on the worst-case convergence quantities in other iteration steps using the lower bound for the next-to-last step. The presented approach can be applied also to other classes of model problems in which the matrix eigenvalues are known, and the Lagrange factors L_k in (2.3) can be evaluated.

Acknowledgments. We thank Miro Rozložník, Daniel Szyld and anonymous referees for helpful comments and suggestions that improved the paper.

Appendix. Let $h = (n+1)^{-1}$, $n \in \mathbb{N}$. Then the following identities hold:

$$\frac{n+1}{2^{2n-1}} \, \frac{1}{\sin^2(k\pi h)} = \prod_{\substack{j=1\\ j\neq k}}^{n} \left| \sin^2\left(\frac{j\pi h}{2}\right) - \sin^2\left(\frac{k\pi h}{2}\right) \right| , \tag{A.1}$$

$$\frac{n+1}{2^n} = \prod_{j=1}^{n} \sin(j\pi h) , \tag{A.2}$$

$$\frac{n}{2} = \sum_{j=1}^{n} \cos^2(j\pi h) = \sum_{j=1}^{n} \sin^2(j\pi h) , \tag{A.3}$$

$$\frac{3\,n-1}{2^3} = \sum_{j=1}^{n} \cos^4\left(\frac{j\pi h}{2}\right) , \tag{A.4}$$

$$\frac{35\,n-29}{2^7} = \sum_{j=1}^{n} \cos^8\left(\frac{j\pi h}{2}\right) , \tag{A.5}$$

ETNA
Kent State University
etna@mcs.kent.edu

$$\frac{n+1}{16} = \sum_{j=1}^{n} \sin^2\left(\frac{j\pi h}{2}\right)\cos^4\left(\frac{j\pi h}{2}\right), \tag{A.6}$$

$$\frac{2\,n(n+2)}{3} = \sum_{j=1}^{n} \sin^{-2}\left(\frac{j\pi h}{2}\right), \tag{A.7}$$

$$\frac{33}{16}\,n - \frac{1}{2} = \sum_{j=1}^{n} \left(\frac{5}{4} - \cos(j\pi h)\right)^2. \tag{A.8}$$

Identity (A.2) can be found in [3, p. 40], and the sums (A.3)–(A.8) can be verified using MAPLE [23]. To prove the non-standard identity (A.1), we note that

$$\begin{aligned}
&\prod_{\substack{j=1\\ j\neq k}}^{n} \left[\sin^2\left(\frac{j\pi h}{2}\right) - \sin^2\left(\frac{k\pi h}{2}\right)\right]\\
&= \prod_{\substack{j=1\\ j\neq k}}^{n} \sin\left(\frac{(j+k)\pi h}{2}\right) \prod_{\substack{j=1\\ j\neq n+1-k}}^{n} \cos\left(\frac{(j+k)\pi h}{2}\right).
\end{aligned}$$

If $kh = \frac{1}{2}$ then, $n+1-k = k$, and the product in (A.1) takes the form

$$\begin{aligned}
\prod_{\substack{j=1\\ j\neq k}}^{n} \left|\sin\left(\frac{(j+k)\pi h}{2}\right)\cos\left(\frac{(j+k)\pi h}{2}\right)\right| &= \frac{1}{2^{n-1}} \prod_{\substack{j=1\\ j\neq k}}^{n} |\sin\left((j+k)\pi h\right)|\\
&= \frac{1}{2^{n-1}} \prod_{j=1}^{n} \sin\left(j\pi h\right) = \frac{n+1}{2^{2n-1}},
\end{aligned}$$

cf. (A.2). Clearly, (A.1) holds since $\sin^2(k\pi h) = 1$ for $kh = \frac{1}{2}$.

If $kh \neq \frac{1}{2}$, then the product in (A.1) can be written as

$$\begin{aligned}
&|\cos(k\pi h)| \prod_{\substack{j=1\\ j\neq k\\ j\neq n+1-k}}^{n} \left|\sin\left(\frac{(j+k)\pi h}{2}\right)\cos\left(\frac{(j+k)\pi h}{2}\right)\right|\\
&= \frac{|\cos(k\pi h)|}{2^{n-2}} \prod_{\substack{j=1\\ j\neq k\\ j\neq n+1-k}}^{n} |\sin\left((j+k)\pi h\right)|\\
&= \frac{|\cos(k\pi h)|}{2^{n-2}|\sin(2k\pi h)|} \cdot \prod_{\substack{j=1\\ j\neq n+1-k}}^{n} |\sin\left((j+k)\pi h\right)|\\
&= \frac{2\sin(k\pi h)\cos(k\pi h)}{2^{n-1}\sin(k\pi h)\sin(2k\pi h)} \cdot \frac{1}{\sin(k\pi h)} \prod_{j=1}^{n} \sin\left(j\pi h\right)\\
&= \frac{n+1}{2^{2n-1}} \frac{1}{\sin^2(k\pi h)}.
\end{aligned}$$

REFERENCES

[1] O. AXELSSON, *Iterative Solution Methods*, Cambridge University Press, 1994.

ETNA
Kent State University
etna@mcs.kent.edu

[2] B. BECKERMANN AND A. B. J. KUIJLAARS, *Superlinear CG convergence for special right-hand sides*, Electron. Trans. Numer. Anal., 14 (2002), pp. 1–19.
[3] I. S. GRADSHTEYN AND I. M. RYZHIK, *Table of integrals, series, and products*, Academic Press Inc., San Diego, CA, sixth ed., 2000. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger.
[4] A. GREENBAUM, *Comparison of splittings used with the conjugate gradient algorithm*, Numer. Math., 33 (1979), pp. 181–193.
[5] ———, *Iterative methods for solving linear systems*, vol. 17 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
[6] A. GREENBAUM AND L. GURVITS, *Max-min properties of matrix factor norms*, SIAM J. Sci. Comput., 15 (1994), pp. 348–358.
[7] A. GREENBAUM AND L. N. TREFETHEN, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comput., 15 (1994), pp. 359–368.
[8] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards , 49 (1952), pp. 409–435.
[9] W. JOUBERT, *A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems*, SIAM J. Sci. Comput., 15 (1994), pp. 427–439.
[10] J. LIESEN AND Z. STRAKOŠ, *Convergence of GMRES for tridiagonal Toeplitz matrices*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 233–251.
[11] J. LIESEN AND P. TICHÝ, *Behavior of CG and MINRES for symmetric tridiagonal Toeplitz matrices*, Preprint 34-2004, Institute of Mathematics, Technical University of Berlin, 2004. Available at http://www.math.tu-berlin.de/preprints.
[12] ———, *Convergence analysis of Krylov subspace methods*, GAMM Mitt. Ges. Angew. Math. Mech., 27 (2004), pp. 153–173 (2005).
[13] ———, *The worst-case GMRES for normal matrices*, BIT, 44 (2004), pp. 79–98.
[14] J. C. MASON AND D. C. HANDSCOMB, *Chebyshev polynomials*, Chapman & Hall/CRC, Boca Raton, FL, 2003.
[15] A. E. NAIMAN, I. M. BABUŠKA, AND H. C. ELMAN, *A note on conjugate gradient convergence*, Numer. Math., 76 (1997), pp. 209–230.
[16] A. E. NAIMAN AND S. ENGELBERG, *A note on conjugate gradient convergence. II, III*, Numer. Math., 85 (2000), pp. 665–683, 685–696.
[17] C. C. PAIGE AND M. A. SAUNDERS, *Solutions of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
[18] Y. SAAD, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, second ed., 2003.
[19] 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.
[20] G. D. SMITH, *Numerical solution of partial differential equations*, The Clarendon Press Oxford University Press, New York, second ed., 1978. Finite difference methods, Oxford Applied Mathematics and Computing Science Series.
[21] THE MATHWORKS, INC., *MATLAB 6.5, Release 13.* Natick, Massachusetts, USA, 2002.
[22] K. TOH, M. TODD, AND R. TÜTÜNCÜ, *SDPT3 – a Matlab software package for semidefinite programming, version 2.1. Interior point methods.* June 2001.
[23] WATERLOO MAPLE, INC., *Maple 8.0.* Waterloo, Ontario, Canada, 2002.