

On Meinardus' Examples For the Conjugate Gradient Method¹

Ren-Cang Li²

January 2005

ABSTRACT

The Conjugate Gradient (CG) method is widely used to solve a positive definite linear system $Ax = b$ of order N . In 1963, Meinardus (*Numer. Math.*, 5 (1963), pp. 14–23.) proved that the relative residual of the k th approximate solution by CG (with the initial approximation $x_0 = 0$) is bounded above by

$$2 \left[\Delta_\kappa^k + \Delta_\kappa^{-k} \right]^{-1} \quad \text{with} \quad \Delta_\kappa = \frac{\sqrt{\kappa} + 1}{\sqrt{\kappa} - 1},$$

where $\kappa \equiv \kappa(A) = \|A\|_2 \|A^{-1}\|_2$ is A 's spectral condition number. In the same paper he also gave an example to achieve this bound for $k = N - 1$ but without saying anything about all other $1 \leq k < N - 1$. It is possible to construct examples to attain Meinardus' bound for any given k , with the help of his example, but such examples depend on k and, furthermore, it will be shown that if the k th residual achieve Meinardus' bound, then the $(k+1)$ th residual must be exactly zero. Therefore it'd be interesting to know if there is any example on which the CG relative residuals are comparable to the bound for all $1 \leq k \leq N - 1$. There are two contributions in this paper.

1. A closed formula for the CG residuals for all $1 \leq k \leq N - 1$ on Meinardus' example is obtained, and in particular it implies that Meinardus' bound is always within a factor of $\sqrt{2}$ of the actual residuals;
2. A complete characterization of extreme positive linear systems for which the k th CG residual achieves Meinardus' bound is also presented. As a consequence, there is no positive linear system whose k th CG residual achieves Meinardus' bound for all $1 \leq k < N$, unless $N = 2$.

¹This report is available on the web at <http://www.ms.uky.edu/~math/MAREport/>.

²Department of Mathematics, University of Kentucky, Lexington, KY 40506 (rccli@ms.uky.edu.) This work was supported in part by the National Science Foundation CAREER award under Grant No. CCR-9875201.

On Meinardus' Examples For the Conjugate Gradient Method

Ren-Cang Li *

June 2005

Abstract

The Conjugate Gradient (CG) method is widely used to solve a positive definite linear system $Ax = b$ of order N . In 1963, Meinardus (*Numer. Math.*, 5 (1963), pp. 14-23.) proved that the relative residual of the k th approximate solution by CG (with the initial approximation $x_0 = 0$) is bounded above by

$$2 [\Delta_\kappa^k + \Delta_\kappa^{-k}]^{-1} \quad \text{with} \quad \Delta_\kappa = \frac{\sqrt{\kappa} + 1}{\sqrt{\kappa} - 1},$$

where $\kappa \equiv \kappa(A) = \|A\|_2 \|A^{-1}\|_2$ is A 's spectral condition number. In the same paper he also gave an example to achieve this bound for $k = N - 1$ but without saying anything about all other $1 \leq k < N - 1$. It is possible to construct examples to attain Meinardus' bound for any given k , with the help of his example, but such examples depend on k and, furthermore, it will be shown that if the k th residual achieve Meinardus' bound, then the $(k + 1)$ th residual must be exactly zero. Therefore it'd be interesting to know if there is any example on which the CG relative residuals are comparable to the bound for all $1 \leq k \leq N - 1$. There are two contributions in this paper.

1. A closed formula for the CG residuals for all $1 \leq k \leq N - 1$ on Meinardus' example is obtained, and in particular it implies that Meinardus' bound is always within a factor of $\sqrt{2}$ of the actual residuals;
2. A complete characterization of extreme positive linear systems for which the k th CG residual achieves Meinardus' bound is also presented. As a consequence, there is no positive linear system whose k th CG residual achieves Meinardus' bound for all $1 \leq k < N$, unless $N = 2$.

1 Introduction

The Conjugate Gradient (CG) method is widely used to solve a positive definite linear system $Ax = b$ (often with certain preconditioning). The basic idea is to seek approximate solutions from the so-called Krylov subspaces. While different implementation may render different numerical behavior, mathematically¹ the k th approximate solution x_k by CG is the optimal

*Department of Mathematics, University of Kentucky, Lexington, KY 40506 (rccli@ms.uky.edu) Supported in part by the National Science Foundation CAREER award under Grant No. CCR-9875201.

¹Without loss of generality, we assume that A is already pre-conditioned and the initial approximation $x_0 = 0$.

one in the sense that [2]

$$\|r_k\|_{A^{-1}} = \min_{x \in \mathcal{K}_k} \|b - Ax\|_{A^{-1}}, \quad (1.1)$$

where $r_k = b - Ax_k$, $\mathcal{K}_k \equiv \mathcal{K}_k(A, b)$ is the k th Krylov subspace of A on b defined as

$$\mathcal{K}_k \equiv \mathcal{K}_k(A, b) \stackrel{\text{def}}{=} \text{span}\{b, Ab, \dots, A^{k-1}b\}, \quad (1.2)$$

and A^{-1} -vector norm $\|z\|_{A^{-1}} \stackrel{\text{def}}{=} \sqrt{z^* A^{-1} z}$. Here the superscript “ $*$ ” takes conjugate transpose. In practice, x_k is computed recursively from x_{k-1} via short term recurrences [2, 3, 5, 12]. But exactly how it is computed, though extremely crucial in practice, is not important to our analysis here in this paper.

CG always converges for positive definite A . In fact, we have the following error bound due to Meinardus [11] (see also [2, 5, 12]):

$$\frac{\|r_k\|_{A^{-1}}}{\|r_0\|_{A^{-1}}} = \min_{x \in \mathcal{K}_k} \frac{\|b - Ax\|_{A^{-1}}}{\|r_0\|_{A^{-1}}} \leq 2 \left[\Delta_\kappa^k + \Delta_\kappa^{-k} \right]^{-1}, \quad (1.3)$$

where $\kappa \equiv \kappa(A) = \|A\|_2 \|A^{-1}\|_2$ is the spectral condition number, generic notation $\|\cdot\|_2$ is for either the spectral norm (the largest singular value) of a matrix or the euclidian length of a vector, and

$$\Delta_t \stackrel{\text{def}}{=} \frac{\sqrt{t} + 1}{|\sqrt{t} - 1|} \quad \text{for } t > 0 \quad (1.4)$$

that will be used frequently later for different t . But how sharp is this bound of Meinardus’? In the same paper [11], Meinardus devised an $N \times N$ positive definite linear system $Ax = b$ for which it was proved that

$$\frac{\|r_{N-1}\|_{A^{-1}}}{\|r_0\|_{A^{-1}}} = 2 \left[\Delta_\kappa^{N-1} + \Delta_\kappa^{-(N-1)} \right]^{-1},$$

but without saying anything about all other $1 \leq k < N - 1$. This example of Meinardus’ (see Remark 2.1 below), can be easily modified to give examples which achieve Meinardus’ bound for any $1 \leq k < N - 1$. This in a sense shows that Meinardus’ bound is sharp and cannot be improved in general. But examples, i.e., A and b , constructed as such depend on the given step-index k and CG on any of these examples for k other than the example was constructed for behaves much differently. So this only proves that Meinardus’ bound is “*locally*” sharp. What about its “*global*” sharpness? I.e.,

Is there any positive definite system $Ax = b$ for which relative residuals $\|r_k\|_{A^{-1}}/\|r_0\|_{A^{-1}}$ achieve Meinardus’ bound for all $1 \leq k < N - 1$?

(1.5)

This question turns out to be too strong and the answer is no (see Theorem 2.3); so instead we ask

Is there any positive definite system $Ax = b$ for which relative residuals $\|r_k\|_{A^{-1}}/\|r_0\|_{A^{-1}}$ are comparable to Meinardus’ bound for all $1 \leq k < N - 1$?

(1.6)

This question has been recently answered positively in Li [7].

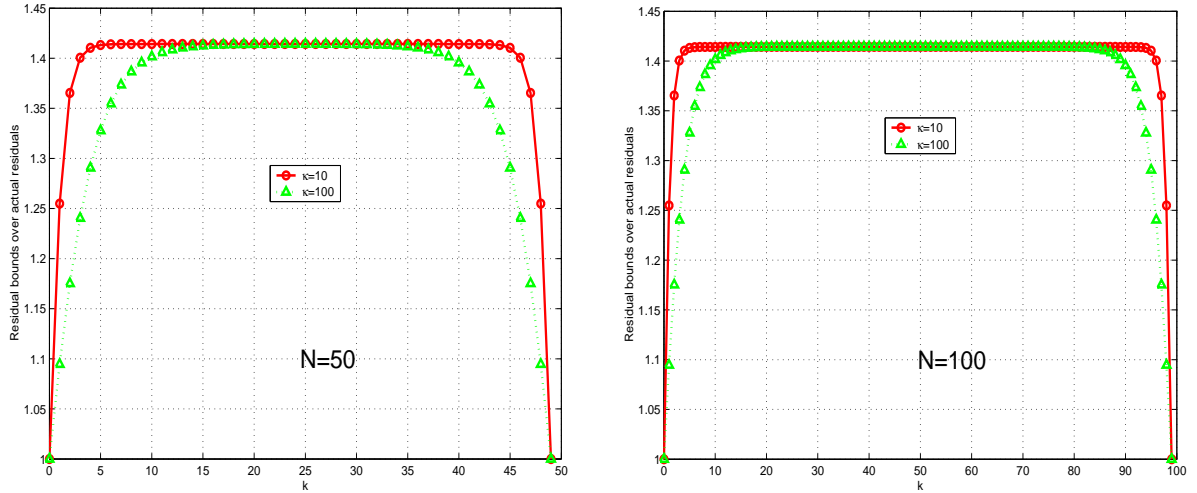


Figure 1.1: Ratios of Meinardus' bound over the exact CG residuals for Meinardus' example.

In this paper, we shall compute the CG residuals on Meinardus' example for all $1 \leq k \leq N - 1$ and investigate extreme positive linear systems for which the k th CG residual achieves Meinardus' bound. Before we set out to do so, let us look at some numerical examples, Figure 1.1 plots the ratios of Meinardus' bound (1.3) over the actual CG relative residuals, i.e., the right-hand side of (1.3) over its left-hand side, on Meinardus' example, where the exact CG residuals were carefully computed within MAPLE² with a sufficient high precision. While it is no surprising at all to see that the ratios are no smaller than 1, they seem to be no bigger than $\sqrt{2}$ as well. This is in fact will be confirmed by one of our main results, which will also furnish another example for the global sharpness question (1.6).

The rest of this paper is organized as follows. Section 2 explains Meinardus' examples and gives our main results – the closed formula for CG residuals for a Meinardus' example and a complete characterizations of extreme positive linear systems for which the k th CG residual achieves Meinardus' bound. Proofs for our main results are rather long and thus given separately in Section 3 and Section 4. Concluding remarks are given in Section 5.

Notation. Throughout this paper, $\mathbb{C}^{n \times m}$ is the set of all $n \times m$ complex matrices, $\mathbb{C}^n = \mathbb{C}^{n \times 1}$, and $\mathbb{C} = \mathbb{C}^1$. Similarly define $\mathbb{R}^{n \times m}$, \mathbb{R}^n , and \mathbb{R} except replacing the word *complex* by *real*. I_n (or simply I if its dimension is clear from the context) is the $n \times n$ identity matrix, and e_j is its j th column. The superscript “ $.T$ ” takes transpose only. We shall also adopt MATLAB-like convention to access the entries of vectors and matrices. $i : j$ is the set of integers from i to j inclusive. For vector u and matrix X , $u_{(j)}$ is u 's j th entry, $X_{(i,j)}$ is X 's (i, j) th entry, $\text{diag}(u)$ is the diagonal matrix with $(\text{diag}(u))_{(j,j)} = u_{(j)}$; X 's submatrices $X_{(k:\ell, i:j)}$, $X_{(k:\ell, :)}$, and $X_{(:, i:j)}$ consist of intersections of row k to row ℓ and column i to column j , row k to row ℓ , and column i to column j , respectively.

²<http://www.maplesoft.com/>.

2 Meinardus' Example and Main Results

The m th Chebyshev polynomial of the 1st kind is

$$T_m(t) = \cos(m \arccos t) \quad \text{for } |t| \leq 1, \quad (2.1)$$

$$= \frac{1}{2} \left(t + \sqrt{t^2 - 1} \right)^m + \frac{1}{2} \left(t - \sqrt{t^2 - 1} \right)^m \quad \text{for } |t| \geq 1. \quad (2.2)$$

It frequently shows up in numerical analysis and computations because of its numerous nice properties, for example $|T_m(t)| \leq 1$ for $|t| \leq 1$ and $|T_m(t)|$ grows extremely fast for $|t| > 1$. It is known (see, e.g., [7])

$$\left| T_m \left(\frac{1+t}{1-t} \right) \right| \equiv \left| T_m \left(\frac{t+1}{t-1} \right) \right| = \frac{1}{2} [\Delta_t^m + \Delta_t^{-m}] \quad \text{for } 1 \neq t > 0. \quad (2.3)$$

$T_m(t)$ has $m + 1$ extreme points in $[-1, 1]$, so-called *the m th Chebyshev extreme nodes*:

$$\tau_{jm} = \cos \vartheta_{jm}, \quad \vartheta_{jm} = \frac{j}{m} \pi, \quad 0 \leq j \leq m, \quad (2.4)$$

at which $|T_m(\tau_{jm})| = 1$. Given $\alpha < \beta$, set

$$\omega = \frac{\beta - \alpha}{2} > 0, \quad \tau = -\frac{\alpha + \beta}{\beta - \alpha}. \quad (2.5)$$

The linear transformation

$$t(z) = \frac{z}{\omega} + \tau = \frac{2}{\beta - \alpha} \left(z - \frac{\alpha + \beta}{2} \right) \quad (2.6)$$

maps $z \in [\alpha, \beta]$ one-to-one and onto $t \in [-1, 1]$. With its inverse transformation $x(t) = \omega(t - \tau)$, we define so-called *the m th translated Chebyshev extreme nodes* on $[\alpha, \beta]$:

$$\tau_{jm}^{\text{tr}} = \omega(\tau_{jm} - \tau), \quad 0 \leq j \leq m. \quad (2.7)$$

It can be verified that $\tau_{0m} = \beta$ and $\tau_{mm} = \alpha$.

Now we are ready to state Meinardus' example. Assume $0 < \alpha < \beta$. Let Q be any $N \times N$ unitary matrix. A Meinardus' example is a positive definite systems $Ax = b$ with

$$A = Q\Lambda Q^*, \quad b = Q\Lambda^{1/2}g, \quad (2.8)$$

where $n = N - 1$, and

$$\Lambda \stackrel{\text{def}}{=} \text{diag}(\tau_{0n}^{\text{tr}}, \tau_{2n}^{\text{tr}}, \dots, \tau_{nn}^{\text{tr}}), \quad g_{(j+1)} \stackrel{\text{def}}{=} \begin{cases} \sqrt{1/\tau_{jn}^{\text{tr}}}, & \text{for } j \in \{0, n\}, \\ \sqrt{2/\tau_{jn}^{\text{tr}}}, & \text{for } 1 \leq j \leq n-1. \end{cases} \quad (2.9)$$

So an example of Meinardus' is any one of them in the family parameterized by unitary Q . Theorem 2.1 is the main result of this paper.

Theorem 2.1 Let $0 < \alpha < \beta$ and let A and b be given by (2.8) and (2.9). r_k is the k th CG residual with initially $r_0 = b$. Then

$$\frac{\|r_k\|_{A^{-1}}}{\|r_0\|_{A^{-1}}} = \rho_k \times 2 \left[\Delta_\kappa^k + \Delta_\kappa^{-k} \right]^{-1} \quad (2.10)$$

for $1 \leq k \leq n$, where $\kappa \equiv \kappa(A) = \beta/\alpha$ and

$$\frac{1}{2} < \frac{1}{2} \left(1 + \frac{2\Delta_\kappa^n}{\Delta_\kappa^{2n} + 1} \right) \leq \rho_k^2 = \frac{1}{2} \left(1 + \frac{\Delta_\kappa^{2k} + \Delta_\kappa^{2(n-k)}}{\Delta_\kappa^{2n} + 1} \right) \leq 1. \quad (2.11)$$

REMARK 2.1 1. As far as the equality is concerned, (2.10) is valid for $k = 0$ as well, which corresponds to the very beginning of CG.

2. The factor ρ_k is symmetrical in k about $n/2$, i.e., $\rho_k = \rho_{n-k}$. This phenomenon certainly showed up in Figure 1.1 which equivalently plotted ρ_k^{-1} .
3. $\rho_k \leq 1$ with equality if and only if $k = 0$ or n .
4. ρ_k is strictly decreasing for $k \leq \lfloor n/2 \rfloor$ (the largest integer that is no bigger than $n/2$) and strictly increasing for $k \geq \lceil n/2 \rceil$ (the smallest integer that is no less than $n/2$), and

$$\frac{1}{\sqrt{2}} < \min_{0 \leq k \leq n} \rho_k = \rho_{\lfloor n/2 \rfloor} \rightarrow \frac{1}{\sqrt{2}} \quad \text{as } n \rightarrow \infty.$$

Theorem 2.1 will be proved through a restatement. It can be verified that the k th CG residual can be reformulated³ as [7]

$$\|r_k\|_{A^{-1}} \equiv \min_{x \in \mathcal{K}_k} \|b - Ax\|_{A^{-1}} = \min_{|u_{(1)}|=1} \|\text{diag}(g) V_{k+1,N}^T u\|_2, \quad (2.12)$$

where, with $\alpha_{j+1} = \tau_{jn}^{\text{tr}}$ for $0 \leq j \leq n$,

$$V_{k+1,N} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_N \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^k & \alpha_2^k & \cdots & \alpha_N^k \end{pmatrix}, \quad (2.13)$$

a $(k+1) \times N$ rectangular Vandermonde matrix. Note also that $\|r_0\|_{A^{-1}} = \|g\|_2$. Therefore Theorem 2.1 can be equivalently stated as follows.

Theorem 2.2 Let $0 < \alpha < \beta$, g as in (2.9), and $V_{k,N}$ as in (2.13) with $\alpha_{j+1} = \tau_{jn}^{\text{tr}}$ for $0 \leq j \leq n$. Then

$$\min_{|u_{(1)}|=1} \frac{\|\text{diag}(g) V_{k,N}^T u\|_2}{\|g\|_2} = \rho_{k-1} \times 2 \left[\Delta_\kappa^{k-1} + \Delta_\kappa^{-(k-1)} \right]^{-1}. \quad (2.14)$$

for $1 \leq k \leq N = n+1$, where $\kappa = \beta/\alpha$.

³A similar reformulation holds for GMRES for normal matrices [6, 10].

$\rho_n = 1$ has already been proved by Meinardus [11]. With it, one can easily construct a positive definite linear system $Ax = b$ for which the k th CG residual achieves Meinardus' bound. For example, A and b are given by (2.8) and (2.9), where $\Lambda = \text{diag}(\tau_{0k}^{\text{tr}}, \tau_{2k}^{\text{tr}}, \dots, \tau_{kk}^{\text{tr}}, \dots)$, i.e., $k + 1$ of A 's eigenvalues are $\tau_{0k}^{\text{tr}}, \tau_{2k}^{\text{tr}}, \dots, \tau_{kk}^{\text{tr}}$, and $g_{(j+1)}$ is $\sqrt{1/\tau_{jk}^{\text{tr}}}$ for $j \in \{0, k\}$ and $\sqrt{2/\tau_{jk}^{\text{tr}}}$ for $1 \leq j \leq k - 1$ and zero for all other j , then $\|r_k\|_{A^{-1}}/\|r_0\|_{A^{-1}} = 2 [\Delta_\kappa^k + \Delta_\kappa^{-k}]^{-1}$. For this example $r_{k+1} = 0$, i.e., convergence occurs at the $(k + 1)$ th step! This is not a coincidence, however. We have the following theorem that characterizes all extreme linear systems as such.

Theorem 2.3 *Let $Ax = b \neq 0$ be a positive definite linear system of order N , and $1 \leq k < N$. If the k th CG residual r_k (initially $r_0 = b$) achieves Meinardus' bound, i.e.,*

$$\frac{\|r_k\|_{A^{-1}}}{\|r_0\|_{A^{-1}}} = 2 [\Delta_\kappa^k + \Delta_\kappa^{-k}]^{-1}, \quad (2.15)$$

where $\kappa \equiv \kappa(A) = \|A\|_2 \|A^{-1}\|_2$, then the following statements hold.

1. $A = Q\Lambda Q^*$ and $b = Q\Lambda^{1/2}g$ for some unitary $Q \in \mathbb{C}^{N \times N}$, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ with $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$, and $g \in \mathbb{R}^N$ with all $g_{(j)} \geq 0$.
2. $\sum_{\lambda_j=\lambda_1} g_{(j)}^2 > 0$ and $\sum_{\lambda_j=\lambda_N} g_{(j)}^2 > 0$.
3. Let $\alpha = \min_j \lambda_j$, and $\beta = \max_j \lambda_j$, and let τ_{jk}^{tr} be the translated Chebyshev extreme nodes on $[\alpha, \beta]$. The distinct λ_j 's in $\{\lambda_j : g_{(j)} > 0\}$ consist of exactly τ_{jk}^{tr} , $0 \leq j \leq k$, i.e.,

$$\{\tau_{jk}^{\text{tr}}, 0 \leq j \leq k\} \subset \{\lambda_j : g_{(j)} > 0\}, \quad \text{and } \lambda_i \in \{\tau_{jk}^{\text{tr}}, 0 \leq j \leq k\} \text{ if } g_{(i)} > 0.$$

4. $r_{k+1} \equiv 0!$

5. Let $\mathbb{J}_\ell = \{j : \lambda_j = \tau_{\ell k}^{\text{tr}}, g_{(j)} > 0\}$. For some constant $\mu > 0$,

$$\|g_{\mathbb{J}_\ell}\|_2 = \mu \begin{cases} \sqrt{1/\tau_{\ell k}^{\text{tr}}}, & \text{for } \ell \in \{0, k\}, \\ \sqrt{2/\tau_{\ell k}^{\text{tr}}}, & \text{for } 1 \leq \ell \leq k - 1. \end{cases} \quad (2.16)$$

As far as CG is concerned, roughly speaking this theorem implies that if the k th CG residual r_k (initially $r_0 = b$) achieves Meinardus' bound, then $Ax = b$ is essentially equivalent to an example of Meinardus' (2.8) and (2.9) with $N = k + 1$. It also implies that unless $N = 2$, there is no positive linear system whose k th CG residual achieves Meinardus' bound for all $1 \leq k < N$.

3 Proof of Theorem 2.2

We will adopt in whole the notation introduced in Section 2 and assume $0 < \alpha < \beta$. Recall, in particular, $n = N - 1$ and A is $N \times N$.

Notice that $T_j(t(z)) \equiv T_j(z/\omega + \tau)$ is a polynomial of degree j in z ; so we write

$$T_j(z/\omega + \tau) = a_{jj}z^j + a_{j-1j}z^{j-1} + \dots + a_{1j}z + a_{0j},$$

where $a_{ij} \equiv a_{ij}(\omega, \tau)$ are functions of ω and τ in (2.5). Their explicit dependence on ω and τ is often suppressed for convenience. For integer $m \geq 1$, define upper triangular $R_m \in \mathbb{R}^{m \times m}$, a matrix-valued function in ω and τ , as

$$R_m \equiv R_m(\omega, \tau) \stackrel{\text{def}}{=} \begin{pmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0m-1} \\ & a_{11} & a_{12} & \cdots & a_{1m-1} \\ & & a_{22} & \cdots & a_{2m-1} \\ & & & \ddots & \vdots \\ & & & & a_{m-1m-1} \end{pmatrix}, \quad (3.1)$$

i.e., the j th column consists of the coefficients of $T_{j-1}(z/\omega + \tau)$. Write $V_N = V_{N,N}$ for short and set

$$\mathbf{S} \stackrel{\text{def}}{=} \begin{pmatrix} T_0(\tau_{0n}) & T_1(\tau_{0n}) & T_2(\tau_{0n}) & \cdots & T_n(\tau_{0n}) \\ T_0(\tau_{1n}) & T_1(\tau_{1n}) & T_2(\tau_{1n}) & \cdots & T_n(\tau_{1n}) \\ \vdots & \vdots & \vdots & & \vdots \\ T_0(\tau_{nn}) & T_1(\tau_{nn}) & T_2(\tau_{nn}) & \cdots & T_n(\tau_{nn}) \end{pmatrix}. \quad (3.2)$$

Then $V_N^T R_N = \mathbf{S}$. Since R_N is upper triangular, we have

$$V_{k,N}^T = (\mathbf{S})_{(:,1:k)} R_k^{-1}, \quad (3.3)$$

a key decomposition of $V_{k,N}^T$ that will play a vital role later in our proofs. Set

$$\Omega = \text{diag}(2^{-1}, 1, 1, \dots, 1, 2^{-1}) \in \mathbb{R}^{N \times N}, \quad \Upsilon \stackrel{\text{def}}{=} \mathbf{S}^T \Omega \mathbf{S}. \quad (3.4)$$

Lemma 3.2 below says Υ is diagonal. So essentially (3.3) gives a QR-like decomposition of $V_{k,N}^T$.

Lemma 3.1 *Let $\vartheta_{kn} = \pi k/n$ as in (2.4). Then*

$$\sum_{k=0}^n \cos \ell \vartheta_{kn} = \begin{cases} N, & \text{if } \ell = 2mn \text{ for some integer } m, \\ 0, & \text{if } \ell \text{ is odd,} \\ 1, & \text{if } \ell \text{ is even, but } \ell \neq 2mn \text{ for any integer } m. \end{cases} \quad (3.5)$$

Proof: Since $\ell \vartheta_{kn} = (\ell k/n)\pi$, the case $\ell = 2mn$ is clear. Assume that $\ell \neq 2mn$ for any integer m , and then $\cos \phi \neq 1$, where $\phi = \ell\pi/n$. Denote $\iota = \sqrt{-1}$. We have

$$\begin{aligned} 2 \sum_{k=0}^n \cos \ell \vartheta_{kn} &= 2 \sum_{k=0}^n \cos k\phi = \sum_{k=0}^n [e^{\iota k\phi} + e^{-\iota k\phi}] \\ &= \sum_{k=0}^n [e^{\iota\phi}]^k + \sum_{k=0}^n [e^{-\iota\phi}]^k \\ &= \frac{1 - [e^{\iota\phi}]^{n+1}}{1 - e^{\iota\phi}} + \frac{1 - [e^{-\iota\phi}]^{n+1}}{1 - e^{-\iota\phi}} \\ &= \frac{1 - e^{\iota(n+1)\phi}}{1 - e^{\iota\phi}} + \frac{1 - e^{-\iota(n+1)\phi}}{1 - e^{-\iota\phi}} \\ &= \frac{1 + \cos n\phi - \cos \phi - \cos(n+1)\phi}{1 - \cos \phi} \\ &= 1 + (-1)^\ell \end{aligned}$$

upon noticing $\cos n\phi = \cos \ell\pi = (-1)^\ell$ and $\cos(n+1)\phi = \cos(\ell\pi + \phi) = (-1)^\ell \cos \phi$. (3.5) is proved. \blacksquare

Lemma 3.2 $\Upsilon = \frac{n}{2}\Omega^{-1}$.

Proof: We notice $(\mathbf{S})_{(i+1,j+1)} = T_j(\tau_{in}) = \cos j\vartheta_{in} = \cos \frac{ji}{n}\pi$, and therefore for $0 \leq i, j \leq n$

$$\begin{aligned}
(\mathbf{S}^T \Omega \mathbf{S})_{(i+1,j+1)} &= \sum_{k=0}^n {}''(\mathbf{S}^T)_{(i+1,k+1)} (\mathbf{S})_{(k+1,j+1)} \\
&= -\frac{T_i(\tau_{0n})T_j(\tau_{0n}) + T_i(\tau_{nn})T_j(\tau_{nn})}{2} + \sum_{k=0}^n T_i(\tau_{kn})T_j(\tau_{kn}) \\
&= -\frac{1 + (-1)^{i+j}}{2} + \sum_{k=0}^n \cos i\vartheta_{kn} \cos j\vartheta_{kn} \\
&= -\frac{1 + (-1)^{i+j}}{2} + \frac{1}{2} \sum_{k=0}^n \cos(i+j)\vartheta_{kn} + \frac{1}{2} \sum_{k=0}^n \cos(i-j)\vartheta_{kn}, \tag{3.6}
\end{aligned}$$

where \sum_j'' means the first and last terms are halved. In Lemma 3.1, let $\ell = i \pm j$, where $0 \leq i, j \leq n$. Since $0 \leq i+j \leq 2n$ and $-n \leq i-j \leq n$, for some integer m

$$\begin{aligned}
i+j = 2mn &\Leftrightarrow i=j=0, \quad \text{or} \quad i=j=n; \\
i-j = 2mn &\Leftrightarrow i=j.
\end{aligned}$$

It follows from (3.6) that $\Upsilon = \frac{n}{2}\Omega^{-1}$. \blacksquare

Lemma 3.3 Let $\Gamma = \text{diag}(\mu + \nu \cos \vartheta_{0n}, \mu + \nu \cos \vartheta_{1n}, \dots, \mu + \nu \cos \vartheta_{nn})$ and define $\Upsilon_{\mu,\nu} \stackrel{\text{def}}{=} \mathbf{S}^T \Omega \Gamma \mathbf{S}$, where $\mu, \nu \in \mathbb{C}$. We have

$$\Upsilon_{\mu,\nu} = \frac{n}{4}\Omega^{-1} (2\mu\Omega + \nu H) \Omega^{-1}, \tag{3.7}$$

where

$$H = \begin{pmatrix} 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & 1 & \ddots & \ddots & \\ & & \ddots & 0 & 1 \\ & & & 1 & 0 \end{pmatrix} \in \mathbb{R}^{N \times N}.$$

Proof: Notice that $\Upsilon_{\mu,\nu} = \mu\Upsilon_{1,0} + \nu\Upsilon_{0,1}$ and $\Upsilon_{1,0} = \mathbf{S}^T \Omega \mathbf{S} = \Upsilon = \frac{n}{2}\Omega^{-1}$ previously defined in (3.4). It is enough to calculate $\Upsilon_{0,1}$. For $0 \leq i, j \leq n$,

$$\begin{aligned}
(\mathbf{S}^T \Omega \Gamma \mathbf{S})_{(i+1,j+1)} &= \sum_{k=0}^n {}''(\mathbf{S}^T)_{(i+1,k)} (\mu + \nu \cos \vartheta_{kn}) (\mathbf{S})_{(k,j+1)} \\
&= \sum_{k=0}^n {}''T_i(\tau_{kn}) (\mu + \nu \cos \vartheta_{kn}) T_j(\tau_{kn})
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^n \cos i\vartheta_{kn} (\mu + \nu \cos \vartheta_{kn}) \cos j\vartheta_{kn} \\
&= \mu \sum_{k=0}^n \cos i\vartheta_{kn} \cos j\vartheta_{kn} + \nu \sum_{k=0}^n \cos i\vartheta_{kn} \cos \vartheta_{kn} \cos j\vartheta_{kn}, \quad (3.8)
\end{aligned}$$

So $(\Upsilon_{0,1})_{(i+1,j+1)} = \sum_{k=0}^n \cos i\vartheta_{kn} \cos \vartheta_{kn} \cos j\vartheta_{kn}$. Now

$$\begin{aligned}
&4 \sum_{k=0}^n \cos i\vartheta_{kn} \cos \vartheta_{kn} \cos j\vartheta_{kn} \\
&= \sum_{k=0}^n \cos(i+j+1)\vartheta_{kn} + \sum_{k=0}^n \cos(i+j-1)\vartheta_{kn} \\
&\quad + \sum_{k=0}^n \cos(i-j+1)\vartheta_{kn} + \sum_{k=0}^n \cos(i-j-1)\vartheta_{kn}. \quad (3.9)
\end{aligned}$$

Apply Lemma 3.1 to conclude $\Upsilon_{0,1} = \frac{n}{4} \Omega^{-1} H \Omega^{-1}$ whose verification is straightforward, albeit tedious. \blacksquare

Lemma 3.4 *Let $m \leq n$ and $\xi \in \mathbb{C}$ such that $(-2\xi\Omega + H)_{(1:m,1:m)}$ is nonsingular. Then the first entry of the solution to $(-2\xi\Omega + H)_{(1:m,1:m)} y = e_1$ is*

$$y_{(1)} = \frac{\gamma_-^m - \gamma_+^m}{\sqrt{\xi^2 - 1}(\gamma_-^m + \gamma_+^m)},$$

where $\gamma_{\pm} = \xi \pm \sqrt{\xi^2 - 1}$.

Proof: Expand y to a 0th entry $y_{(0)}$ and a $(m+1)$ th entry $y_{(m+1)}$ satisfying

$$y_{(0)} - \xi y_{(1)} = -1, \quad y_{(m+1)} = 0. \quad (3.10)$$

Entry-wise, we have

$$y_{(i-1)} - 2\xi y_{(i)} + y_{(i+1)} = 0, \quad \text{for } 1 \leq i \leq m.$$

The general solution has form $y_{(i)} = c_+ \gamma_+^i + c_- \gamma_-^i$, where γ_{\pm} are the two roots of $1 - 2\xi\gamma + \gamma^2 = 0$, i.e., $\gamma_{\pm} = \xi \pm \sqrt{\xi^2 - 1}$. We now determine c_+ and c_- by the edge conditions (3.10):

$$\begin{aligned}
(1 - \xi\gamma_+) c_+ + (1 - \xi\gamma_-) c_- &= -1, \\
\gamma_+^{m+1} c_+ + \gamma_-^{m+1} c_- &= 0.
\end{aligned}$$

Notice $\gamma_+ \gamma_- = 1$ and

$$\begin{aligned}
(1 - \xi\gamma_+) \gamma_-^{m+1} - (1 - \xi\gamma_-) \gamma_+^{m+1} &= (\gamma_- - \xi) \gamma_-^m - (\gamma_+ - \xi) \gamma_+^m \\
&= -\sqrt{\xi^2 - 1} (\gamma_-^m + \gamma_+^m)
\end{aligned}$$

to get

$$c_+ = \frac{-\gamma_-^{m+1}}{-\sqrt{\xi^2 - 1}(\gamma_-^m + \gamma_+^m)}, \quad c_- = \frac{+\gamma_+^{m+1}}{-\sqrt{\xi^2 - 1}(\gamma_-^m + \gamma_+^m)}.$$

Finally $y_{(1)} = c_+ \gamma_+ + c_- \gamma_-$. \blacksquare

In its present form, the next lemma was proved in [7]. But it was also implied by the proof of [6, Theorem 2.1]. See also [9].

Lemma 3.5 *If Z has full column rank. Then*

$$\min_{|u_{(1)}|=1} \|Zu\|_2 = [e_1^T (Z^*Z)^{-1} e_1]^{-1/2}. \quad (3.11)$$

Proof of Theorem 2.2. By Lemma 3.5,

$$\min_{|u_{(1)}|=1} \frac{\|\text{diag}(g)V_{k,N}^T u\|_2}{\|g\|_2} = \frac{\left[e_1^T \left(V_{k,N} [\text{diag}(g)]^2 V_{k,N}^T \right)^{-1} e_1 \right]^{-1/2}}{\|g\|_2}. \quad (3.12)$$

Let $\Gamma = \text{diag}(\tau_{0n}^{\text{tr}}, \tau_{1n}^{\text{tr}}, \dots, \tau_{nn}^{\text{tr}}) \equiv \text{diag}(\mu + \nu \cos \vartheta_{00}, \mu + \nu \cos \vartheta_{01}, \dots, \mu + \nu \cos \vartheta_{nn})$, where $\mu = -\omega\tau$ and $\nu = \omega$ as in (2.5). Then

$$\begin{aligned} V_{k,N} [\text{diag}(g)]^2 V_{k,N}^T &= 2V_{k,N} \Gamma^{-1} \Omega V_{k,N}^T \\ &= 2 \begin{pmatrix} e^T \\ V_{k-1,N} \Gamma \end{pmatrix} \Gamma^{-1} \Omega \begin{pmatrix} e & \Gamma V_{k-1,N}^T \end{pmatrix} \\ &= 2 \begin{pmatrix} e^T \Gamma^{-1} \Omega e & e^T \Omega V_{k-1,N}^T \\ V_{k-1,N} \Omega e & V_{k-1,N} \Gamma \Omega V_{k-1,N}^T \end{pmatrix}, \end{aligned} \quad (3.13)$$

where $e = (1, 1, \dots, 1)^T$. Notice $V_{k-1,N}^T = (\mathbf{S})_{(:,1:k-1)} R_{k-1}^{-1}$ by (3.3) to get

$$\begin{aligned} V_{k-1,N} \Omega e &= V_{k-1,N} \Omega V_{k-1,N}^T e_1 \\ &= R_{k-1}^{-*} (\Upsilon_{1,0})_{(1:k-1,1:k-1)} R_{k-1}^{-1} e_1 \\ &= R_{k-1}^{-*} (\Upsilon_{1,0})_{(1:k-1,1:k-1)} e_1, \end{aligned} \quad (3.14)$$

$$\begin{aligned} V_{k-1,N} \Gamma \Omega V_{k-1,N}^T &= R_{k-1}^{-*} [(\mathbf{S})_{(:,1:k-1)}]^T \Gamma \Omega (\mathbf{S})_{(:,1:k-1)} R_{k-1}^{-1} \\ &= R_{k-1}^{-*} (\Upsilon_{\mu,\nu})_{(1:k-1,1:k-1)} R_{k-1}^{-1}, \end{aligned} \quad (3.15)$$

in the notation introduced in Lemma 3.3. Recall (see, e.g., [13, Page 23]),

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}^{-1} = \begin{pmatrix} C_{11}^{-1} & -C_{11}^{-1} B_{12} B_{22}^{-1} \\ -B_{22}^{-1} B_{21} C_{11}^{-1} & B_{22}^{-1} + B_{22}^{-1} B_{21} C_{11}^{-1} B_{12} B_{22}^{-1} \end{pmatrix},$$

assuming all inversions exist, where $C_{11} = B_{11} - B_{12} B_{22}^{-1} B_{21}$. We have from (3.13)

$$\begin{aligned} &e_1^T (V_{k,N} [\text{diag}(g)]^2 V_{k,N}^T)^{-1} e_1 \\ &= \frac{1}{2} \left[\zeta - e^T \Omega V_{k-1,N}^T (V_{k-1,N} \Gamma \Omega V_{k-1,N}^T)^{-1} V_{k-1,N} \Omega e \right]^{-1}, \end{aligned} \quad (3.16)$$

where $\zeta = e^T \Gamma^{-1} \Omega e$. But, from (3.14) and (3.15),

$$\begin{aligned} &e^T \Omega V_{k-1,N}^T (V_{k-1,N} \Gamma \Omega V_{k-1,N}^T)^{-1} V_{k-1,N} \Omega e \\ &= e_1^T (\Upsilon_{1,0})_{(1:k-1,1:k-1)} \left[(\Upsilon_{\mu,\nu})_{(1:k-1,1:k-1)} \right]^{-1} (\Upsilon_{1,0})_{(1:k-1,1:k-1)} e_1 \\ &= n^2 e_1^T \left[(\Upsilon_{\mu,\nu})_{(1:k-1,1:k-1)} \right]^{-1} e_1, \end{aligned} \quad (3.17)$$

and for $k \leq N$, by Lemma 3.4 with $m = k - 1$ and $\xi = \tau$,

$$\begin{aligned} e_1^T \left[(\Upsilon_{\mu, \nu})_{(1:k-1, 1:k-1)} \right]^{-1} e_1 &= n^{-1} e_1^T \left[(2\mu\Omega + \nu H)_{(1:k-1, 1:k-1)} \right]^{-1} e_1 \\ &= \frac{1}{n\omega} e_1^T \left[(-2\tau\Omega + H)_{(1:k-1, 1:k-1)} \right]^{-1} e_1 \\ &= \frac{1}{n\omega} \frac{\gamma_-^{k-1} - \gamma_+^{k-1}}{\sqrt{\tau^2 - 1}(\gamma_-^{k-1} + \gamma_+^{k-1})}, \end{aligned} \quad (3.18)$$

where $\gamma_{\pm} = \tau \pm \sqrt{\tau^2 - 1}$. The conditions of Lemma 3.4 are fulfilled because $|\tau| > 1$ and $-2\tau\Omega + H$ is diagonally dominant and thus nonsingular. Since $2\zeta = \|g\|_2^2$, we have by (3.12) and (3.16) – (3.18)

$$\min_{|u_{(1)}|=1} \frac{\|\text{diag}(g)V_{k,N}^T u\|_2}{\|g\|_2} = \left[1 - \frac{n}{\omega\zeta\sqrt{\tau^2 - 1}} \frac{\gamma_-^{k-1} - \gamma_+^{k-1}}{\gamma_-^{k-1} + \gamma_+^{k-1}} \right]^{1/2}. \quad (3.19)$$

We now compute $\omega\zeta\sqrt{\tau^2 - 1}$. Let $f(z) \stackrel{\text{def}}{=} \prod_{j=0}^n (z - \tau_{jn}^{\text{tr}})$. Then

$$f(z) = \eta(z - \tau_{0n}^{\text{tr}})(z - \tau_{nn}^{\text{tr}})U_{n-1}(z/\omega + \tau)$$

for some constant η , where $U_{n-1}(t)$ is the $(n-1)$ th Chebyshev polynomial of the second kind. This is because the zeros of $U_{n-1}(z/\omega + \tau)$ are precisely $\tau_{jn}^{\text{tr}} = \omega(\tau_{jn} - \tau)$, $j = 1, 2, \dots, n-1$. Then, upon noticing $\tau_{0n}^{\text{tr}} = \beta$ and $\tau_{nn}^{\text{tr}} = \alpha$,

$$\begin{aligned} 2\zeta &= \sum_{j=0}^n \frac{2}{\tau_{jn}^{\text{tr}}} = -\frac{1}{\tau_{0n}^{\text{tr}}} + 2 \sum_{j=0}^n \frac{1}{\tau_{jn}^{\text{tr}}} - \frac{1}{\tau_{nn}^{\text{tr}}} = -\left(\frac{1}{\alpha} + \frac{1}{\beta}\right) - 2\frac{f'(0)}{f(0)} \\ &= -\frac{\alpha + \beta}{\alpha\beta} - 2\frac{-(\alpha + \beta)U_{n-1}(\tau) + \alpha + \beta U'_{n-1}(\tau)/\omega}{\alpha\beta U_{n-1}(\tau)} \\ &= \frac{\alpha + \beta}{\alpha\beta} - \frac{2U'_{n-1}(\tau)}{\omega U_{n-1}(\tau)}. \end{aligned}$$

Recall [1, Page 37]

$$2U_{n-1}(t) = \frac{(t + \sqrt{t^2 - 1})^n - (t - \sqrt{t^2 - 1})^n}{\sqrt{t^2 - 1}}, \quad (3.20)$$

$$2U'_{n-1}(t) = n \frac{(t + \sqrt{t^2 - 1})^n + (t - \sqrt{t^2 - 1})^n}{t^2 - 1} - \frac{t \left[(t + \sqrt{t^2 - 1})^n - (t - \sqrt{t^2 - 1})^n \right]}{(t^2 - 1)\sqrt{t^2 - 1}}.$$

They yield

$$2U_{n-1}(\tau) = \frac{\gamma_+^n - \gamma_-^n}{\sqrt{\tau^2 - 1}}, \quad 2U'_{n-1}(\tau) = n \frac{\gamma_+^n + \gamma_-^n}{\tau^2 - 1} - \frac{\tau(\gamma_+^n - \gamma_-^n)}{(\tau^2 - 1)\sqrt{\tau^2 - 1}}.$$

Therefore, upon noticing $\omega = (\alpha + \beta)/2$ and $\tau = -(\beta + \alpha)/(\beta - \alpha)$,

$$\begin{aligned} 2\zeta &= \frac{\alpha + \beta}{\alpha\beta} + \frac{2}{\omega} \frac{n}{\sqrt{\tau^2 - 1}} \frac{\gamma_-^n + \gamma_+^n}{\gamma_-^n - \gamma_+^n} + \frac{2}{\omega} \frac{\tau}{\tau^2 - 1} \\ &= \frac{2}{\omega} \frac{n}{\sqrt{\tau^2 - 1}} \frac{\gamma_-^n + \gamma_+^n}{\gamma_-^n - \gamma_+^n}, \end{aligned} \quad (3.21)$$

$$\omega\zeta\sqrt{\tau^2 - 1} = n \frac{\gamma_-^n + \gamma_+^n}{\gamma_-^n - \gamma_+^n}. \quad (3.22)$$

Equation (3.19) and (3.22) imply

$$\min_{|u_{(1)}|=1} \frac{\|\text{diag}(g)V_{k,N}^T u\|_2}{\|g\|_2} = \left[1 - \frac{\gamma_-^n - \gamma_+^n \gamma_-^{k-1} - \gamma_+^{k-1}}{\gamma_-^n + \gamma_+^n \gamma_-^{k-1} + \gamma_+^{k-1}} \right]^{1/2}. \quad (3.23)$$

Because $\tau = -(\kappa + 1)/(\kappa - 1)$,

$$\gamma_-^{k-1} = (-1)^{k-1} \Delta_\kappa^{k-1}, \quad \gamma_+^{k-1} = (-1)^{k-1} \Delta_\kappa^{-(k-1)},$$

and therefore

$$\min_{|u_{(1)}|=1} \frac{\|\text{diag}(g)V_{k,N}^T u\|_2}{\|g\|_2} = \left[1 - \frac{\Delta_\kappa^n - \Delta_\kappa^{-n} \Delta_\kappa^{k-1} - \Delta_\kappa^{-(k-1)}}{\Delta_\kappa^n + \Delta_\kappa^{-n} \Delta_\kappa^{k-1} + \Delta_\kappa^{-(k-1)}} \right]^{1/2}. \quad (3.24)$$

For $k = N \equiv n + 1$, the right-hand side of (3.24) is $2[\Delta_\kappa^n + \Delta_\kappa^{-n}]^{-1}$, as was shown by Meinardus [11]. For any other k , we have

$$\min_{|u_{(1)}|=1} \frac{\|\text{diag}(g)V_{k,N}^T u\|_2}{\|g\|_2} = \rho_{k-1} \times 2 \left[\Delta_\kappa^{k-1} + \Delta_\kappa^{-(k-1)} \right]^{-1}. \quad (3.25)$$

where

$$\begin{aligned} \rho_{k-1} &\stackrel{\text{def}}{=} \frac{\text{RHS of (3.24)}}{2 \left[\Delta_\kappa^{k-1} + \Delta_\kappa^{-(k-1)} \right]^{-1}} \\ &= \left[\frac{\left(\Delta_\kappa^{k-1} + \Delta_\kappa^{-(k-1)} \right)^2}{4} - \frac{(\Delta_\kappa^n - \Delta_\kappa^{-n}) \left(\Delta_\kappa^{2(k-1)} - \Delta_\kappa^{-2(k-1)} \right)}{4 \left(\Delta_\kappa^n + \Delta_\kappa^{-n} \right)} \right]^{1/2} \\ &= \left[\frac{1}{2} \frac{\left(\Delta_\kappa^{k-1} + \Delta_\kappa^{-(k-1)} \right) \left(\Delta_\kappa^{n-(k-1)} + \Delta_\kappa^{-[n-(k-1)]} \right)}{\Delta_\kappa^n + \Delta_\kappa^{-n}} \right]^{1/2} \\ &= \left[\frac{1}{2} \frac{\left(\Delta_\kappa^{2(k-1)} + 1 \right) \left(\Delta_\kappa^{2[n-(k-1)]} + 1 \right)}{\Delta_\kappa^{2n} + 1} \right]^{1/2} \end{aligned}$$

which yields (2.11). ■

4 Proof of Theorem 2.3

We first prove two general lemmas for Vandermonde matrix $V_N \equiv V_{N,N}$ as defined in (2.13) with arbitrary, possibly complex, nodes α_j .

Lemma 4.1 *Assume one or more of 1) there are less than n distinct α_j , 2) some $\alpha_j = 0$, and 3) some $g_{(j)} = 0$ occur. Then*

$$\min_{|u_{(1)}|=1} \|\text{diag}(g)V_N^T u\|_2 = \begin{cases} 0, & \text{if all } \alpha_j \neq 0; \\ \sqrt{\sum_{\alpha_j=0} |g_{(j)}|^2}, & \text{otherwise.} \end{cases}$$

Proof: If all $\alpha_j \neq 0$, only Case 1) and 3) are possible. Let ℓ be the number of distinct α_j 's, exclude those corresponding to $g_{(j)} = 0$. Then $\ell < n$. By permutating the rows of $\text{diag}(g)V_N^T$, we may assume that $\alpha_1, \alpha_2, \dots, \alpha_\ell$ are distinct and for α_j ($j > \ell$) either it is equal to some α_i ($i \leq \ell$) or corresponding $g_{(j)} = 0$. Set $v \in \mathbb{C}^N$ whose $v_{(j)}$ is the coefficient of z^{j-1} in the polynomial $\phi(z) = \prod_{j=1}^\ell (z - \alpha_j)$. $v_{(1)} = \prod_{j=1}^\ell (-\alpha_j) \neq 0$. We have

$$\min_{|u_{(1)}|=1} \|\text{diag}(g)V_N^T u\|_2 \leq \|\text{diag}(g)V_N^T (v/v_{(1)})\|_2 = 0,$$

as expected.

If some $\alpha_j = 0$. Since $\|\text{diag}(g)V_N^T u\|_2 \geq \sqrt{\sum_{\alpha_j=0} |g_{(j)}|^2}$ always for any vector u with $|u_{(1)}| = 1$, it suffices to find a vector u to annihilate all other rows corresponding to $\alpha_j \neq 0$. Such u can be constructed similarly to what we just did. ■

Lemma 4.2 *Let $V_N \equiv V_{N,N}$ be as defined in (2.13) with all nodes α_j (possibly complex) distinct, and let $f(z) = \prod_{j=1}^N (z - \alpha_j)$.*

1. *If all $g_{(j)} \neq 0$, then*

$$\min_{|u_{(1)}|=1} \frac{\|\text{diag}(g)V_N^T u\|_2}{\|g\|_2} = \left[\sqrt{\sum_{j=1}^N \left(\frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right)^2 |g_{(j)}|^{-2}} \sqrt{\sum_{j=1}^N |g_{(j)}|^2} \right]^{-1}. \quad (4.1)$$

2.

$$\max_g \min_{|u_{(1)}|=1} \frac{\|\text{diag}(g)V_N^T u\|_2}{\|g\|_2} = \left[\sqrt{\sum_{j=1}^N \frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|}} \right]^{-1}, \quad (4.2)$$

where the maximum is achieved if and only if for some constant $\mu > 0$,

$$|g_{(j)}| = \mu \left[\frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right]^{1/2} \quad \text{for } 1 \leq j \leq N. \quad (4.3)$$

Proof: This lemma is essentially [10, Theorems 2.1 and 3.1], but stated differently. The proof below has a slightly different flavor. In Lemma 3.5, take $Z = \text{diag}(g)V_N^T$. The assumptions make this Z nonsingular. Therefore

$$\begin{aligned} \left[\min_{|u_{(1)}|=1} \|\text{diag}(g)V_N^T u\|_2 \right]^{-2} &= e_1^T (\bar{V}_N \Phi V_N^T)^{-1} e_1 \\ &= e_1^T (V_N \Phi V_N^*)^{-1} e_1 \\ &= (V_N^{-1} e_1)^* \Phi^{-1} (V_N^{-1} e_1), \end{aligned}$$

where $\Phi = [\text{diag}(g)]^* \text{diag}(g)$ and \bar{V}_N is the complex conjugate of V_N . Let $y = V_N^{-1} e_1$, the first column of V_N^{-1} which consists of the constant terms of the Lagrangian basis functions:

$$\ell_j(z) = \prod_{i \neq j} \frac{z - \alpha_i}{\alpha_i - \alpha_j}, \quad 1 \leq j \leq N,$$

since $\ell_j(\alpha_i) = 1$ for $i = j$ and 0 otherwise, which means the j th row of V_N^{-1} consists of the coefficients of $\ell_j(z)$. Therefore

$$\begin{aligned} e_1^T (\bar{V}_N \Phi V_N^T)^{-1} e_1 &= \sum_{j=1}^N \left(\frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right)^2 |g_{(j)}|^{-2}, \\ \min_{|u_{(1)}|=1} \frac{\|\text{diag}(g) V_N^T u\|_2}{\|g\|_2} &= \left[\sqrt{\sum_{j=1}^N \left(\frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right)^2 |g_{(j)}|^{-2}} \sqrt{\sum_{j=1}^N |g_{(j)}|^2} \right]^{-1} \\ &\leq \left[\sqrt{\sum_{j=1}^N \frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|}} \right]^{-1}, \end{aligned} \quad (4.4)$$

where it is an equality if and only if $|g_{(j)}|$ are given by (4.3). \blacksquare

REMARK 4.1 This lemma closely relates to a result of Greenbaum [4, (2.2) and Theorem 1] which in our notation essentially proved that *if all nodes $\alpha_j > 0$, there exist k of α_j 's: $\alpha_{j_1}, \dots, \alpha_{j_k}$ such that*

$$\max_g \min_{|u_{(1)}|=1} \frac{\|\text{diag}(g) V_{k,N}^T u\|_2}{\|g\|_2} = \max_h \min_{|u_{(1)}|=1} \frac{\|\text{diag}(h) V_k^T u\|_2}{\|h\|_2},$$

where V_k is the $k \times k$ Vandermonde matrix with nodes α_{j_i} . Notice the difference in conditions: Lemma 4.3 only covers $k = N$, while this result of Greenbaum's is for all $1 \leq k \leq N$ but requires all $\alpha_j > 0$. Greenbaum [4, Theorem 1] also obtained an expression for the optimal h but a bit of more complicated than we get from applying Lemma 4.3. It is not clear how to find out the most relevant nodes α_{j_i} .

Lemma 4.3 *Let $\omega, \tau \in \mathbb{C}$ (not necessarily associated with any interval $[\alpha, \beta]$ as previously required), and let $n = N - 1$ and $\tau_{j_n}^{\text{tr}}$ as in (2.7) with any given ω and τ . Suppose Vandermonde matrix V_N has nodes $\alpha_{j+1} = \tau_{j_n}^{\text{tr}}$ for $0 \leq j \leq n$.*

1. *If all $g_{(j)} \neq 0$, then*

$$\begin{aligned} \min_{|u_{(1)}|=1} \frac{\|\text{diag}(g) V_N^T u\|_2}{\|g\|_2} &= \frac{n\omega}{|\tau_{0n}^{\text{tr}} \tau_{nn}^{\text{tr}} U_{n-1}(\tau)|} \times \\ &\left[\sqrt{\frac{1}{(2\tau_{0n}^{\text{tr}})^2} |g_{(1)}|^{-2} + \sum_{j=2}^{N-1} \frac{1}{(\tau_{j_n}^{\text{tr}})^2} |g_{(j+1)}|^{-2} + \frac{1}{(2\tau_{nn}^{\text{tr}})^2} |g_{(N)}|^{-2}} \sqrt{\sum_{j=1}^N |g_{(j)}|^2} \right]^{-1}, \end{aligned} \quad (4.5)$$

where $U_{n-1}(t)$ is the $(n - 1)$ th Chebyshev polynomial of the second kind as in (3.20).

2.

$$\max_g \min_{|u_{(1)}|=1} \frac{\|\text{diag}(g) V_N^T u\|_2}{\|g\|_2} = \frac{n\omega}{|\tau_{0n}^{\text{tr}} \tau_{nn}^{\text{tr}} U_{n-1}(\tau)|} \left[\frac{1}{|2\tau_{0n}^{\text{tr}}|} + \sum_{j=1}^{n-1} \frac{1}{|\tau_{j_n}^{\text{tr}}|} + \frac{1}{|2\tau_{nn}^{\text{tr}}|} \right]^{-1}, \quad (4.6)$$

where the maximum is achieved if and only if for some $\mu > 0$

$$|g_{(j+1)}| = \begin{cases} \mu \sqrt{1/|\tau_{jn}^{\text{tr}}|}, & \text{for } j \in \{0, n\}, \\ \mu \sqrt{2/|\tau_{jn}^{\text{tr}}|}, & \text{for } 1 \leq j \leq n-1. \end{cases} \quad (4.7)$$

Proof: $f(z) = \prod_{j=1}^N (z - \alpha_j)$ admits

$$f(z) = \eta (z - \tau_{0n}^{\text{tr}})(z - \tau_{nn}^{\text{tr}})U_{n-1}(z/\omega + \tau),$$

where η^{-1} is the coefficient of z^{n-1} in $U_{n-1}(z/\omega + \tau)$. We have

$$\begin{aligned} f(0) &= \eta \tau_{0n}^{\text{tr}} \tau_{nn}^{\text{tr}} U_{n-1}(\tau), \\ f'(\tau_{0n}^{\text{tr}}) &= -\eta (\tau_{0n}^{\text{tr}} - \tau_{nn}^{\text{tr}}) U_{n-1}(1) \\ &= -\eta (\tau_{0n}^{\text{tr}} - \tau_{nn}^{\text{tr}}) n \\ &= -\eta 2n\omega, \\ f'(\tau_{nn}^{\text{tr}}) &= -\eta (\tau_{nn}^{\text{tr}} - \tau_{0n}^{\text{tr}}) U_{n-1}(-1) \\ &= (-1)^n \eta (\tau_{nn}^{\text{tr}} - \tau_{0n}^{\text{tr}}) n \\ &= -(-1)^n \eta 2n\omega, \end{aligned}$$

and for $1 \leq j \leq n-1$

$$\begin{aligned} f'(\tau_{jn}^{\text{tr}}) &= \eta (\tau_{jn}^{\text{tr}} - \tau_{0n}^{\text{tr}})(\tau_{jn}^{\text{tr}} - \tau_{nn}^{\text{tr}})U'_{n-1}(\tau_{jn})/\omega \\ &= \eta (\tau_{jn}^{\text{tr}} - \tau_{0n}^{\text{tr}})(\tau_{jn}^{\text{tr}} - \tau_{nn}^{\text{tr}})n/[\omega(1 - \tau_{jn}^2)] \\ &= -\eta n\omega. \end{aligned}$$

Therefore by Lemma 4.2, we have (4.5) and (4.6). \blacksquare

REMARK 4.2 As a corollary to (4.6) and Meinardus' bound, we deduce that the right-hand side of (4.6) is equal to $|T_n(\tau)| = 2[\Delta_\kappa^n + \Delta_\kappa^{-n}]^{-1}$.

Proof of Theorem 2.3. Item 1 is always true for any given positive definite system $Ax = b$. In fact let $A = \tilde{Q}\Lambda\tilde{Q}$ be its eigendecomposition, where \tilde{Q} is unitary, and Λ as in the theorem since A is positive definite. Set $\tilde{g} = \Lambda^{-1/2}\tilde{Q}^*b$. Define $g = (|\tilde{g}_{(1)}|, |\tilde{g}_{(2)}|, \dots, |\tilde{g}_{(N)}|)^T \in \mathbb{R}^N$. Then $\tilde{g} = Dg$ for some diagonal D with $|D_{(j,j)}| = 1$. Finally $A = Q\Lambda Q^*$ and $b = Q\Lambda^{1/2}g$ with $Q = \tilde{Q}D$ still unitary.

Next we notice that

$$\begin{aligned} \|r_k\|_{A^{-1}} &= \min_{x \in \mathcal{K}_k} \|b - Ax\|_{A^{-1}} = \min_{p_k(0)=1} \|p_k(\Lambda)g\|_2 \\ &= \min_{p_k(0)=1} \sqrt{\sum_{j=1}^N |p_k(\lambda_j)|^2 g_{(j)}^2}, \end{aligned} \quad (4.8)$$

where $p_k(z)$ denotes a polynomial of degree no more than k . If either inequality in Item 2 is violated, the effective condition number $\kappa' < \kappa(A)$ as far as CG is concerned and the Meinardus' bound gives

$$\frac{\|r_k\|_{A^{-1}}}{\|r_0\|_{A^{-1}}} \leq 2 \left[\Delta_{\kappa'}^k + \Delta_{\kappa'}^{-k} \right]^{-1} < 2 \left[\Delta_\kappa^k + \Delta_\kappa^{-k} \right]^{-1},$$

contradicting (2.15). This proves Item 2.

For Item 3, we first claim that λ_j for which $g_{(j)} > 0$ is in $\{\tau_{jk}^{\text{tr}}, 0 \leq j \leq k\}$. Otherwise if there was a j_0 such that $g_{(j_0)} > 0$ and $\lambda_{j_0} \notin \{\tau_{jk}^{\text{tr}}, 0 \leq j \leq k\}$, then $|T_k(\lambda_{j_0}/\omega + \tau)| < 1$, where ω and τ are given by (2.5). Now take $p_k(z) = q_k(z)$ in (4.8), where $q_k(z) = T_k(z/\omega + \tau)/T_k(\tau)$, to get

$$\begin{aligned} \|r_k\|_{A^{-1}} &\leq \sqrt{|q_k(\lambda_{j_0})|^2 g_{(j_0)}^2 + \sum_{j \neq j_0} |q_k(\lambda_j)|^2 g_{(j)}^2} \\ &< |T_k(\tau)|^{-1} \sqrt{g_{(j_0)}^2 + \sum_{j \neq j_0} g_{(j)}^2} \\ &= |T_k(\tau)|^{-1} \|r_0\|_{A^{-1}}, \end{aligned}$$

contradicting (2.15). This proves the claim. On the other hand, since $r_k \neq 0$, there are at least $k+1$ distinct values in $\{\lambda_j : g_{(j)} > 0\}$ and therefore $\{\lambda_j : g_{(j)} > 0\} \supset \{\tau_{jk}^{\text{tr}}, 0 \leq j \leq k\}$. Item 3 is proved.

Item 3 says effectively A has $k+1$ distinct eigenvalues as far as CG is concerned and thus $r_{k+1} = 0$. This is Item 4.

Define $\hat{g} \in \mathbb{R}^{k+1}$ by $\hat{g}_{(\ell+1)} = \|g_{\mathbb{J}_\ell}\|_2$. (4.8) gives

$$\|r_k\|_{A^{-1}} = \min_{p_k(0)=1} \sqrt{\sum_{j=0}^k |p_k(\tau_{jk}^{\text{tr}})|^2 \hat{g}_{(j+1)}^2} = \min_{|u_{(1)}|=1} \|\text{diag}(\hat{g})V_{k+1}^T\|_2,$$

where $V_{k+1} \equiv V_{k+1, k+1}$ is the $(k+1) \times (k+1)$ Vandermonde matrix as defined in (2.13) with nodes $\alpha_{j+1} = \tau_{jk}^{\text{tr}}$ for $0 \leq j \leq k$. The condition (2.15) and Meinardus' bound (1.3) implies that for \hat{g}

$$\min_{|u_{(1)}|=1} \|\text{diag}(\hat{g})V_{k+1}^T\|_2 = \max_h \min_{|u_{(1)}|=1} \|\text{diag}(h)V_{k+1}^T\|_2.$$

Lemma 4.3 shows $\hat{g}_{(\ell+1)} = \|g_{\mathbb{J}_\ell}\|_2$ must take the form of (2.16). ■

5 Concluding remarks

We have found a closed formula for the CG residuals for Meinardus' examples. These residuals may deviate from the well-known Meinardus' bounds by a factor no bigger than $1/\sqrt{2}$, indicating Meinardus' bounds governing CG convergence rate is very tight in general. Three key technical components that made our computations possible are

1. transforming CG residual computations as minimization problems involving rectangular Vandermonde matrices,
2. the QR-like decomposition $V_N = \mathbf{S}R_N^{-1}$, and
3. the solution to $\min_{|u_{(1)}|=1} \|Zu\|_2$.

It turns out that QR-like decompositions exist for quite a few Vandermonde matrices, and the combination of the three technical components have been used in [7, 8] for arriving at the

asymptotically optimally conditioned real Vandermonde matrices, analyzing the sharpness of existing error bounds for CG and the symmetric Lanczos method for eigenvalue problems.

We completely characterized the extreme positive linear systems for which the k th CG residuals achieve Meinardus' bound. Roughly speaking, as far as CG is concerned, these extreme examples are nothing but a Meinardus' example of order $k + 1$. As a consequence, unless $N = 2$ there is no positive linear system whose k th CG residual achieves Meinardus' bound for all $1 \leq k < N$.

References

- [1] P. BORWEIN AND T. ERDÉLYI, *Polynomials and Polynomial Inequalities*, vol. 161 of Graduate Texts in Mathematics, Springer, New York, 1995.
- [2] J. DEMMEL, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
- [3] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, 3rd ed., 1996.
- [4] A. GREENBAUM, *Comparison of splittings used with the conjugate gradient algorithm*, Numer. Math., 33 (1979), pp. 181–194.
- [5] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.
- [6] I. C. F. IPSEN, *Expressions and bounds for the GMRES residual*, BIT, 40 (2000), pp. 524–535.
- [7] R.-C. LI, *Sharpness in rates of convergence for CG and symmetric Lanczos methods*, Technical Report 2005-01, Department of Mathematics, University of Kentucky, 2005. Available at <http://www.ms.uky.edu/~math/MAREport/>.
- [8] ———, *Vandermonde matrices with Chebyshev nodes*, Technical Report 2005-02, Department of Mathematics, University of Kentucky, 2005. Available at <http://www.ms.uky.edu/~math/MAREport/>.
- [9] J. LIESEN, M. ROZLOZNÍK, AND Z. STRAKOS, *Least squares residuals and minimal residual methods*, SIAM J. Sci. Comput., 23 (2002), pp. 1503–1525.
- [10] J. LIESEN AND P. TICHÝ, *The worst-case GMRES for normal matrices*, BIT, 44 (2004), pp. 79–98.
- [11] G. MEINARDUS, *Über eine Verallgemeinerung einer Ungleichung von L. V. Kantorowitsch*, Numer. Math., 5 (1963), pp. 14–23.
- [12] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2nd ed., 2003.
- [13] K. ZHOU, J. C. DOYLE, AND K. GLOVER, *Robust and Optimal Control*, Prentice Hall, 1995.