ON MEINARDUS’ EXAMPLES FOR THE CONJUGATE GRADIENT METHOD

REN-CANG LI

Abstract. The Conjugate Gradient (CG) method is widely used to solve a positive definite linear system $Ax = b$ of order $N$. It is well-known that the relative residual of the $k$th approximate solution by CG (with the initial approximation $x_0 = 0$) is bounded above by

$$2 \left[ \Delta_k^k + \Delta_k^{-k} \right]^{-1}$$

with $\Delta_k = \frac{\sqrt{\kappa} + 1}{\sqrt{\kappa} - 1}$,

where $\kappa \equiv \kappa(A) = \|A\|_2 \|A^{-1}\|_2$ is $A$’s spectral condition number. In 1963, Meinardus (Numer. Math., 5 (1963), pp. 14-23) gave an example to achieve this bound for $k = N - 1$ but without saying anything about all other $1 \leq k < N - 1$. This very example can be used to show that the bound is sharp for any given $k$ by constructing examples to attain the bound, but such examples depend on $k$ and for them the $(k+1)$th residual is exactly zero. Therefore it would 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 the 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 the bound is also presented.

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, mathematically the $k$th approximate solution $x_k$ by CG is the optimal one in the sense that the $k$th approximation error $A^{-1}b - x_k$ satisfies [11, Theorem 6.1]

$$\|A^{-1}b - x_k\|_A = \min_{x \in \mathcal{K}_k} \|A^{-1}b - x\|_A,$$

or, equivalently, the $k$th residual $r_k = b - Ax_k$ satisfies

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

Date: Received September 2005, Revised January 2006.
1991 Mathematics Subject Classification. 65F10.
Keywords and phrases. Conjugate gradient method, Krylov subspace, rate of convergence, Vandermonde matrix, condition number.

This work was supported in part by the National Science Foundation CAREER award under Grant No. CCR-9875201 and by the National Science Foundation under Grant No. DMS-0510664.

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

1
where $K_k \equiv K_k(A, b)$ is the $k$th Krylov subspace of $A$ on $b$ defined as

$$K_k \equiv K_k(A, b) \triangleq \text{span}\{b, Ab, \ldots , A^{k-1}b\},$$

and $M$-vector norm $\|z\|_M \triangleq \sqrt{z^* M z}$. Here the superscript “$*$” takes conjugate transpose. In practice, $x_k$ is computed recursively from $x_{k-1}$ via short term recurrences [4, 7, 10, 19]. 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 well-known and frequently referenced error bound (see, e.g., [4, 10, 19, 22]):

$$\|r_k\|_{A^{-1}} \leq \|A^{-1}b - x_k\|_A \leq 2 \left[ \Delta_k + \Delta_k^{-1} \right]^{-1},$$

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

$$\Delta_t \triangleq \frac{\sqrt{t+1}}{\sqrt{t-1}} \quad \text{for} \; t > 0$$

that will be used frequently later for different $t$. The widely cited Kaniel [13] (1966) gave a proof of (1.4) while saying “This result is known” with a pointer to Meinardus [18] (1963). The same bound was proved to hold for Richardson-like processes [5, pp.28-31], making it likely that (1.4) could be known before 1963 because their proofs do not differ much.

This paper is concerned with how sharp this well-known error bound is. Consider $A = \text{diag}(\lambda_1, \lambda_2, \ldots , \lambda_n)$ and $b$ the vector of all ones, where $\lambda_j \in [\delta, 1]$ either randomly or equidistantly distributed on the interval. Figure 1.1 plots the ratios of the residual bounds by (1.4) over the actual residuals. What it shows is that initially for small $k$, bounds by (1.4) are good indications of actual residuals, but as $k$ becomes larger and larger, this bound overestimates the actual ones too much.
to be of any use. Are the phenomena in Figure 1.1 representative? Often this is what people observed [20], known as superlinear convergence.

In [18], Meinardus devised an \( N \times N \) positive definite linear system \( Ax = b \) for which he proved that

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

but without saying anything about all other \( 1 \leq k < N - 1 \). This example of Meinardus’ can be easily modified to give examples which achieve the error bound in (1.4) for any given \( 1 \leq k < N - 1 \), e.g., trivially embed an example of Meinardus’ of dimension \( k + 1 \) with zero blocks to make it of dimension \( N \). Greenbaum [10, p.52] and [9] claimed that the error bound for a given \( k \) can be attained for \( A \) having eigenvalues at the extreme points of a translated Chebyshev polynomial of degree \( k \) and some particular \( b \) whose explicit form was not given, however. The same conclusion can be reached from [1, p.561] if \( A \)’s eigenvalues are chosen as Greenbaum’s. This in a sense shows that the error bound in (1.4) 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 and in particular \( r_{k+1} = 0 \) exactly. So this only proves that the error bound is “locally” sharp: for given \( k \),

\[
\max_{\kappa(A)=\gamma} \frac{2 \left[ \Delta_k^{k} + \Delta_{\gamma}^{\gamma-k} \right]^{-1}}{\|r_k\|_A^{-1}/\|r_0\|_A^{-1}} = 1.
\]

What about its “global” sharpness? I.e.,

\[
\text{Is there any positive definite system } Ax = b \text{ for which relative residuals } \frac{\|r_k\|_A^{-1}}{\|r_0\|_A^{-1}} \text{ achieve the error bounds by (1.4) for all } 1 \leq k < N - 1? 
\]

This question turns out to be too strong and the answer is no by Kaniel [13, Theorem 4.4] who showed that if \( r_k \) attains the bound, then \( r_{k+1} = 0 \), (see also Theorem 2.2 below); so instead we ask

\[
\text{Is}
\]

\[
\sup_{\kappa(A)=\gamma} \max_{1 \leq k \leq n-1} \frac{2 \left[ \Delta_k^{k} + \Delta_{\gamma}^{\gamma-k} \right]^{-1}}{\|r_k\|_A^{-1}/\|r_0\|_A^{-1}} \text{ modestly bounded?}
\]

This question has been recently answered positively in Li [14], using \( A \) with eigenvalues being the translated zeros of \( N \)th Chebyshev polynomial of the first kind. It is proved there that the ratio in (1.8) can be bounded from above by a bound that asymptotically approaches to \( \sqrt{2} \Delta_{\gamma}/\sqrt{\Delta_{\gamma}} - 1 \) as \( k \) goes to infinity. It depends on \( \kappa(A) = \gamma \) and, unfortunately, can be arbitrarily large as \( \gamma \to 1^+ \). A much stronger bound on the ratio, namely \( \sqrt{2} \), is implied later in this paper.

In what follows, we shall compute the CG residuals on Meinardus’ examples for all \( 1 \leq k \leq N - 1 \) and investigate extreme positive linear systems for which the \( k \)th CG residual achieves the error bound in (1.4). Before we set out to do so, let us look at some numerical examples, Figure 1.2 plots the ratios of the error bounds by (1.4) over the actual CG relative residuals, i.e., the right-hand side of (1.4) over
its left-hand side, on a Meinardus’ example, where the exact CG residuals were carefully computed within MAPLE\(^2\) with a sufficiently high precision. While it is not surprising at all to see that the ratios are not smaller than 1, they seem not to be bigger than \(\sqrt{2}\) as well. This in fact will be confirmed by one of our main results, which will also furnish another example for the global sharpness question (1.8), in addition to the one in [14].

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 characterization of extreme positive linear systems for which the \(k\)th CG residual achieves the error bound in (1.4). 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 “\(\cdot^T\)” 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 \(X(k:\ell,i:j)\), \(X(k:\ell,:j)\), and \(X(:,:,i:j)\) consist of intersections of row \(k\) to row \(\ell\) and column \(i\) to column \(j\), row \(k\) to row \(\ell\), and column \(i\) to column \(j\), respectively.

2. **Meinardus’ Examples and Main Results**

The \(m\)th Chebyshev polynomial of the 1st kind is

\[
T_m(t) = \cos(m \arccos t) \quad \text{for } |t| \leq 1, \\
= \frac{1}{2} \left( t + \sqrt{t^2 - 1} \right)^m + \frac{1}{2} \left( t - \sqrt{t^2 - 1} \right)^m \quad \text{for } |t| \geq 1.
\]

\(^{\text{2http://www.maplesoft.com/}}\)
It frequently shows up in numerical analysis and computations because of its numerous nice properties, for example
\[ |T_m(t)| \leq 1 \text{ for } |t| \leq 1 \text{ and } |T_m(t)| \text{ grows extremely fast for } |t| > 1. \]
Later we will need
\[ \left| T_m \left( \frac{1 + t}{1 - t} \right) \right| = \frac{1}{2} \left[ \Delta^m_r + \Delta^m_{-r} \right] \quad \text{for } 1 \neq t > 0. \]
The first equality holds because \( T_m(-t) = (-1)^m T_m(t) \). We shall prove the 2nd equality for \( 0 < t < 1 \) only and a proof for \( t > 1 \) is similar. For \( 0 < t < 1 \), we have
\[ \left| T_m \left( 1 + \frac{t}{1 - t} \right) \right| = \frac{1 + t + 2\sqrt{t}}{1 - t} = \frac{1 + \sqrt{t}}{1 - \sqrt{t}} = \Delta_t, \]
which proves (2.3) for \( 0 < t < 1 \). \( T_m(t) \) has \( m + 1 \) extreme points in \([-1, 1]\), so-called the \( m \)th Chebyshev extreme nodes:
\[ \tau_{jm} = \cos \vartheta_{jm}, \quad \vartheta_{jm} = \frac{j}{m} \pi, \quad 0 \leq j \leq m, \]
at which \( |T_m(\tau_{jm})| = 1 \). Given \( \alpha < \beta \), set
\[ \omega = \frac{\beta - \alpha}{2} > 0, \quad \tau = -\frac{\alpha + \beta}{\beta - \alpha}. \]
The linear transformation
\[ t(z) = \frac{z}{\omega} + \frac{2}{\beta - \alpha} \left( z - \frac{\alpha + \beta}{2} \right) \]
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 \([\alpha, \beta]\):
\[ \tau_{jm}^t = \omega(\tau_{jm} - \tau), \quad 0 \leq j \leq m. \]
It can be verified that \( \tau_{0m} = \beta \) and \( \tau_{mm} = \alpha \).
Now we are ready to state Meinardus’ examples. Assume \( 0 < \alpha < \beta \). For the sake of presentation, set
\[ n = N - 1. \]
Let \( Q \) be any \( N \times N \) unitary matrix. A Meinardus’ example is a positive definite system \( Ax = b \) with
\[ A = Q \Lambda Q^*, \quad b = Q \Lambda^{1/2} g, \]
where
\[ \Lambda \overset{\text{def}}{=} \text{diag} (\tau_{0m}^t, \tau_{1m}^t, \ldots, \tau_{nm}^t), \quad g(j+1) \overset{\text{def}}{=} \left\{ \begin{array}{ll} \frac{1}{\sqrt{\tau_{jm}^t}}, & \text{for } j \in \{0, n\}, \\ \frac{2}{\sqrt{\tau_{jm}^t}}, & \text{for } 1 \leq j \leq n - 1. \end{array} \right. \]
So an example of Meinardus’ is any member of the family parameterized by unitary \( Q \). Theorem 2.1 is one of the two main results 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
\[ \| r_k \|_{A^{-1}} \| r_0 \|_{A^{-1}} = \rho_k \times 2 \left[ \Delta^k + \Delta^{-k} \right]^{-1} \]
for $1 \leq k \leq n$, where $\kappa \equiv \kappa(A) = \beta/\alpha$ and

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

**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 $\rho_k$ is symmetrical in $k$ about $n/2$, i.e., $\rho_k = \rho_{n-k}$. This phenomenon certainly showed up in Figure 1.2 which equivalently plotted $\rho_k^{-1}$.
3. $\rho_k \leq 1$ with equality if and only if $k = 0$ or $n$.
4. $\rho_k$ 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

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

The fact that $\rho_n = 1$ has already been established by Meinardus [18]. With it, one can easily construct a positive definite linear system $Ax = b$ for which the $k$th CG residual achieves the error bound in (1.4). For example, $A$ and $b$ are given by (2.8) and (2.9), where $\Lambda = \text{diag}(\tau_0^\nu, \tau_1^\nu, \ldots, \tau_{n-1}^\nu, \ldots)$, i.e., $k + 1$ of $A$'s eigenvalues are $\tau_0^\nu, \tau_1^\nu, \ldots, \tau_{n-1}^\nu$, and $g(j+1)$ is $1/\tau_j^\nu$ for $j \in \{0, k\}$ and $\sqrt{2/\tau_j^\nu}$ for $1 \leq j \leq k - 1$ and zero for all other $j$, then $\|r_k\|_A^{-1}/\|r_0\|_A^{-1} = 2 [\Delta_k^2 + \Delta_{n-k}^2]^{-1}$.

For this example $r_{k+1} = 0$, i.e., convergence occurs at the $(k+1)$th step! This is not a coincidence, as it must be due to Kaniel [13, Theorem 4.4]. The following theorem characterizes all extreme linear systems as such.

**Theorem 2.2.** 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 the error bound in (1.4), i.e.,

$$
\frac{\|r_k\|_A^{-1}}{\|r_0\|_A^{-1}} = 2 [\Delta_k^2 + \Delta_{n-k}^2]^{-1},
$$

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

1. $\Lambda = QAQ^*$ and $b = QA^{1/2}g$ for some unitary $Q \in \mathbb{C}^{N \times N}$, $\Lambda = \text{diag} (\lambda_1, \lambda_2, \ldots, \lambda_N)$ with $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$, and $g \in \mathbb{R}^N$ with all $g(j) \geq 0$.
2. $\sum_{j=1}^{\lambda_j} \lambda_j g_j^2 > 0$ and $\sum_{j=\lambda_j}^{n} \lambda_j g_j^2 > 0$.
3. Let $\alpha = \min \lambda_j$, and $\beta = \max \lambda_j$, and let $\tau_j^\nu$ be the translated Chebyshev extreme nodes on $[\alpha, \beta]$. The distinct $\lambda_j$’s in $\{\lambda_j : g(j) > 0\}$ consist of exactly $\tau_j^\nu$, $0 \leq j \leq k$, i.e.,

$$
\{\tau_j^\nu, 0 \leq j \leq k\} \subset \{\lambda_j : g(j) > 0\}, \text{ and } \lambda_i \in \{\tau_j^\nu, 0 \leq j \leq k\} \text{ if } g(i) > 0.
$$

4. [13, Theorem 4.4] $r_{k+1} \equiv 0$.
5. Let $\mathbb{J}_\ell = \{ j : \lambda_j = \tau_{r_j}^\nu, g(j) > 0 \}$. For some constant $\mu > 0$,

$$
\|g_\ell\|_2 = \mu \left\{ \begin{array}{ll}
\sqrt{1/\tau_{r_j}^\nu}, & \text{for } \ell \in \{0, k\}, \\
\sqrt{2/\tau_{r_j}^\nu}, & \text{for } 1 \leq \ell \leq k - 1.
\end{array} \right.
$$
REMARK 2.2. Any \( Ax = b \) described by Items 1, 2, 3, and 5 is essentially equivalent to an example of Meinardus’ (2.8) and (2.9) with \( N = k + 1 \). Therefore this theorem practically says that the \( k \)th CG residual \( r_k \) (initially \( r_0 = b \)) achieves the error bound in (1.4) if and only if \( Ax = b \) is an example of Meinardus’. Thus unless \( N = 2 \), there is no positive linear system whose \( k \)th CG residual achieves the error bound in (1.4) for all \( 1 \leq k < N \).

3. PROOF OF THEOREM 2.1

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 \( N \times N \).

Theorem 2.1 will be proved through a restatement. For \( A \) as in (2.8) and (2.9),

\[
\min_{x \in \mathbb{K}_k} \| b - Ax \|_{A^{-1}} = \min_{\phi_k(0) = 1} \| \phi_k(A)b \|_{A^{-1}} = \min_{\phi_k(0) = 1} \| \phi_k(A)A^{-1/2}Q^*b \|_2 = \min_{\phi_k(0) = 1} \| \phi_k(A)g \|_2 = \min_{|u_{(1)}| = 1} \| \text{diag}(g)V_{k+1,n}u \|_2,
\]

where \( \phi_k(t) \) is a polynomial of degree \( k \), \( u \in \mathbb{C}^{k+1} \), and with \( \alpha_{j+1} = \tau_{jn}^\alpha \) for \( 0 \leq j \leq n \),

\[
V_{k+1,N} \overset{\text{def}}{=} \begin{pmatrix}
1 & 1 & \cdots & 1 \\
\alpha_1 & \alpha_2 & \cdots & \alpha_N \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_1^k & \alpha_2^k & \cdots & \alpha_N^k
\end{pmatrix},
\]

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

Theorem 3.1. Let \( 0 < \alpha < \beta \), \( g \) as in (2.9), and \( V_{k,N} \) as in (3.2) with \( \alpha_{j+1} = \tau_{jn}^\alpha \) for \( 0 \leq j \leq n \). Then

\[
\min_{|u_{(1)}| = 1} \frac{\| \text{diag}(g)V_{k,N}^*u \|_2}{\| g \|_2} = \rho_{k-1} \times 2 \left[ \Delta_k^{k-1} + \Delta_n^{(k-1)} \right]^{-1},
\]

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

The rest of this section is devoted to the proof of this theorem. Notice that \( T_j(t(z)) \equiv T_j(z/\omega + \tau) \) is a polynomial of degree \( j \) in \( z \); so we write

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

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

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

$$S_N \overset{\text{def}}{=} \begin{pmatrix} T_0(\tau_{00}) & T_0(\tau_{10}) & \ldots & T_0(\tau_{nn}) \\ T_1(\tau_{00}) & T_1(\tau_{10}) & \ldots & T_1(\tau_{nn}) \\ \vdots & \vdots & \ddots & \vdots \\ T_n(\tau_{00}) & T_n(\tau_{10}) & \ldots & T_n(\tau_{nn}) \end{pmatrix}. $$

Then $V_N^T R_N = S_N^T$. Since $R_N$ is upper triangular, we have

$$V_{k,N}^T = S_{k,N}^T R_k^{-1}, $$

a key decomposition of $V_{k,N}^T$ that will play a vital role later in our proofs, where $S_{k,N} = (S_N)_{(1,k,:)}$ is $S_N$'s first $k$ rows. Set

$$\Omega = \text{diag}(2^{-1}, 1, 1, \ldots, 1) \in \mathbb{R}^{N \times N}, \quad \Upsilon \overset{\text{def}}{=} S_N \Omega S_N^T. $$

Lemma 3.2 below says $\Upsilon$ is diagonal. So essentially (3.6) gives a QR-like decomposition of $V_{k,N}^T$.

Lemma 3.1 below is probably well-known but a precise reference is hard to find. However, it can be proved by using either Euler identity $\cos \theta = (e^{i\theta} + e^{-i\theta})/2$, where $i = \sqrt{-1}$, the imaginary unit, or the identities in [8, p.30]. Detail is omitted.

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

$$\sum_{k=0}^{n} \cos \ell \vartheta_{kn} = \begin{cases} N, & \text{if } \ell = 2mn \text{ for some integer } m, \\ 0, & \text{if } \ell \text{ is odd}, \\ 1, & \text{if } \ell \text{ is even, but } \ell \neq 2mn \text{ for any integer } m. \end{cases} \quad (3.8)$$

Lemma 3.2 is known to [2, p.33], and probably long before that. A proof can be given with the help of Lemma 3.1, and again detail is omitted.

**Lemma 3.2.** Let $S_N$, $\Omega$, and $\Upsilon$ be defined as in (3.5) and (3.7). Then $\Upsilon = \frac{N}{4} \Omega^{-1}$. 

**Lemma 3.3.** Let $\Gamma = \text{diag}(\mu + \nu \cos \vartheta_{00}, \mu + \nu \cos \vartheta_{10}, \ldots, \mu + \nu \cos \vartheta_{nn})$ and define $\Upsilon_{\mu,\nu} \overset{\text{def}}{=} S_N \Omega \Gamma S_N^T$, where $\mu, \nu \in \mathbb{C}$. We have

$$\Upsilon_{\mu,\nu} = \frac{N}{4} \Omega^{-1} (2\mu \Omega + \nu H) \Omega^{-1}, \quad (3.9)$$

where

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

**Proof.** Notice that $\Upsilon_{\mu,\nu} = \mu \Upsilon_{1,0} + \nu \Upsilon_{0,1}$ and $\Upsilon_{1,0} = S_N \Omega S_N^T = \Upsilon = \frac{N}{4} \Omega^{-1}$ by Lemma 3.2. It is enough to calculate $\Upsilon_{0,1}$. Using $\sum''$ to mean the first and last
terms halved, we have for $0 \leq i, j \leq n$,

$$
(3.10) \quad (S_N \Omega S_N^T)_{(i+1,j+1)} = \sum_{k=0}^{n}'' (S_N)_{(i+1,k)} \left( \mu + \nu \cos \vartheta_{kn} \right) \left( S_N^T \right)_{(k,j+1)}
$$

$$
= \sum_{k=0}^{n}'' T_j(\tau_{kn}) \left( \mu + \nu \cos \vartheta_{kn} \right) \cos j\vartheta_{kn}
$$

$$
= \mu \sum_{k=0}^{n}'' \cos i\vartheta_{kn} \cos j\vartheta_{kn}
$$

$$
+ \nu \sum_{k=0}^{n}'' \cos i\vartheta_{kn} \cos \vartheta_{kn} \cos j\vartheta_{kn}.
$$

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

$$
4 \sum_{k=0}^{n}'' \cos i\vartheta_{kn} \cos \vartheta_{kn} \cos j\vartheta_{kn}
$$

$$
= \sum_{k=0}^{n}'' \cos(i+j+1)\vartheta_{kn} + \sum_{k=0}^{n}'' \cos(i+j-1)\vartheta_{kn}
$$

$$
+ \sum_{k=0}^{n}'' \cos(i-j+1)\vartheta_{kn} + \sum_{k=0}^{n}'' \cos(i-j-1)\vartheta_{kn}.
$$

Apply Lemma 3.1 to conclude $\Upsilon_{0,1} = \frac{2}{\xi} \Omega^{-1} H \Omega^{-1}$ whose verification is straightforward, albeit tedious.

**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 have a 0th entry $y_{(0)}$ and a $(m+1)$th entry $y_{(m+1)}$ satisfying

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

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 $y_{(i)} = c_+ \gamma_+^{m+1} + c_- \gamma_-^{m+1}$, where $\gamma_{\pm}$ 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 (3.11):

$$
(1 - \xi \gamma_+) c_+ + (1 - \xi \gamma_-) c_- = -1, \quad \gamma_+^{m+1} c_+ + \gamma_-^{m+1} c_- = 0.
$$
Notice $\gamma_+ \gamma_- = 1$ and 

$$(1 - \xi \gamma_+) \gamma_-^{-1} - (1 - \xi \gamma_-) \gamma_+^{-1} = (\gamma_- - \xi \gamma_-) \gamma_+^{-1} - (\gamma_- - \xi) \gamma_-$$

$$= -\sqrt{\xi^2 - 1} (\gamma_- \gamma_+)$$

to get 

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

Finally $y(1) = c_+ \gamma_+ + c_- \gamma_-$. $\square$

In its present general form, the next lemma was proved in [14]. It was also implied by the proof of [12, Theorem 2.1]. See also [16].

**Lemma 3.5.** If $Z$ has full column rank, then 

$$(3.12) \quad \min_{|u(1)|=1} \|Zu\|_2 = |e_1^T (Z^* Z)^{-1} e_1|^{-1/2}. $$

**Proof.** Set $v = Zu$. Since $Z$ has full column rank, its Moore-Penrose pseudo-inverse is $Z^* = (Z^* Z)^{-1} Z^*$ [21] and thus $u = Z^* v$. This gives a one-one and onto mapping between $u \in \mathbb{C}^m$ and the column space $v \in \text{span}(Z)$. Now 

$$(3.13) \quad \min_{|u(1)|=1} \|Zu\|_2 = \min_u \frac{\|Zu\|_2}{|u(1)|} = \min_{v \in \text{span}(Z)} \frac{\|v\|_2}{|e_1^T Z^* v|} \geq \min_v \frac{\|v\|_2}{|e_1^T Z^* v|} = \|e_1^T Z^*\|_2^{-1}, $$

where the last min is achieved at 

$$v_{\text{opt}} = (e_1^T Z^* )^* = Z(Z^* Z)^{-1} e_1 \in \text{span}(Z)$$

which implies the “$\geq$” in (3.13) is actually an equality, and $u_{\text{opt}} = Z^* v_{\text{opt}} / |e_1^T Z^* v_{\text{opt}}|$. Finally 

$$\|e_1^T Z^*\|_2 = \sqrt{|e_1^T Z^* (Z^*)^* e_1|} = \sqrt{|e_1^T (Z Z^*)^{-1} e_1|}. $$

This completes the proof. $\square$

**Proof of Theorem 3.1.** By Lemma 3.5,

$$(3.14) \quad \min_{|u(1)|=1} \|\text{diag}(g) V_{k,N}^T u\|_2 = \frac{|e_1^T \left(V_{k,N} \text{diag}(g)^2 V_{k,N}^T\right)^{-1} e_1|^{-1/2}}{\|g\|_2}. $$

Let $\Gamma = \text{diag}(\tau_{00}, \tau_{11}, \ldots, \tau_{nn}) \equiv \text{diag}(\mu + \nu \cos \theta_{00}, \mu + \nu \cos \theta_{01}, \ldots, \mu + \nu \cos \theta_{nn})$, where $\mu = -\omega \tau$ and $\nu = \omega$ as in (2.5). Then 

$$(3.15) \quad V_{k,N} \text{diag}(g)^2 V_{k,N}^T = 2 V_{k,N} \Gamma^{-1} \Omega V_{k,N}^T$$

$$= 2 \left( V_{k-1,N} \Gamma\right)^{-1} \Omega \left( V_{k-1,N}^T \right)$$

$$= 2 \left( V_{k-1,N} \Omega e \quad e^T \Omega V_{k-1,N}^T \right).$$
where \( e = (1, 1, \ldots, 1)^T \). Notice \( V_{k-1,N}^T = S^T_{k-1,N}R^{-1}_{k-1} \) by (3.6) to get

\[
V_{k-1,N} \Omega e = V_{k-1,N} \Omega V_{k-1,N}^T e_1 = R^{-1}_{k-1} (\mathcal{T}_{1,0})_{(1,k-1,k-1)} R^{-1}_{k-1} e_1 = R^{-1}_{k-1} (\mathcal{T}_{1,0})_{(1,k-1,k-1)} e_1,
\]

(3.16)

and for \( \zeta \)

\[
V_{k-1,N} \Gamma \Omega V_{k-1,N}^T = R^{-1}_{k-1} S_{k-1,N} \Gamma \Omega S^T_{k-1,N} R^{-1}_{k-1} = R^{-1}_{k-1} (\mathcal{T}_{\mu,\nu})_{(1,k-1,k-1)} R^{-1}_{k-1},
\]

(3.17)

in the notation introduced in Lemma 3.3. Recall\(^\text{3}\)

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

assuming all inversions exist, where \( C_{11} = B_{11} - B_{12} B_{22}^{-1} B_{21} \). We have from (3.15)

\[
e_1^T (V_{k,N} \text{diag}(\mathcal{G}))^2 V_{k,N}^T e_1 = \frac{1}{2} \left[ \zeta - e^T \Gamma^{-1} \Omega \left( V_{k-1,N} \Gamma \Omega V_{k-1,N}^T \right)^{-1} V_{k-1,N} \Omega e \right]^{-1},
\]

where \( \zeta = e^T \Gamma^{-1} \Omega e \). But, from (3.16) and (3.17),

\[
e_1^T \left( \mathcal{T}_{1,0} \right)_{(1,k-1,k-1)} \left( \mathcal{T}_{\mu,\nu} \right)_{(1,k-1,k-1)}^{-1} \left( \mathcal{T}_{1,0} \right)_{(1,k-1,k-1)} e_1 = n^2 e_1^T \left( \mathcal{T}_{\mu,\nu} \right)_{(1,k-1,k-1)}^{-1} e_1,
\]

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

\[
e_1^T \left( \mathcal{T}_{\mu,\nu} \right)_{(1,k-1,k-1)}^{-1} e_1 = \frac{1}{2} \left[ (2 \mu \Omega + \nu H)_{(1,k-1,k-1)} \right]^{-1} e_1 = \frac{1}{\nu \omega} \left[ (-2 \tau \Omega + H)_{(1,k-1,k-1)} \right]^{-1} e_1 = \frac{1}{\nu \omega} \frac{1}{\sqrt{\tau^2 - 1}} \frac{1}{\gamma_+ - \gamma_-}.
\]

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 (3.14) and (3.18) – (3.20)

\[
\min_{\|u\|_1 = 1} \frac{\|\text{diag}(\mathcal{G}) V_{k,N}^T u\|_2}{\|g\|_2} = \left[ 1 - \frac{n}{\omega \zeta \sqrt{\tau^2 - 1} (\gamma_+ - \gamma_-)} \right]^{1/2}.
\]

(3.21)

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

\[
f(z) = \eta \prod_{j=0}^n (z - \tau_{j,n}^\nu) U_{n-1}(z/\omega + \tau)
\]

\(^3\)This is well-known. See, e.g., [6, pp.102-103], [23, Page 23].
for some constant $\eta$, where $U_{n-1}(t)$ is the $(n - 1)$th Chebyshev polynomial of the second kind. This is because the zeros of $U_{n-1}(\omega + \tau)$ are precisely $\tau_{jn}^n = \omega(\tau_{jn} - \tau), j = 1, 2, \ldots, n - 1$. Then, upon noticing $\tau_{in}^n = \beta$ and $\tau_{mn}^n = \alpha,$

$$
2\zeta = \sum_{j=0}^{n} \frac{2}{\tau_{jn}^n} = -\frac{1}{\tau_{on}^n} + 2 \sum_{j=0}^{n} \frac{1}{\tau_{jn}^n} = -\left(\frac{1}{\alpha} + \frac{1}{\beta}\right) - 2\frac{f'(0)}{f(0)}
$$

$$
= -\frac{\alpha + \beta}{\alpha\beta} - 2\frac{(\alpha + \beta)U_{n-1}(\tau) + \alpha + \beta U'_{n-1}(\tau)/\omega}{\alpha\beta U_{n-1}(\tau)}
$$

$$
= \frac{\alpha + \beta}{\alpha\beta} - \frac{2 U'_{n-1}(\tau)}{\omega U_{n-1}(\tau)}.
$$

Recall [3, Page 37]

$(3.22)$

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

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

$(t^2 - 1)\sqrt{t^2 - 1}.$

They yield

$$
2U_{n-1}(\tau) = \frac{\gamma^n_+ - \gamma^n_-}{\sqrt{\tau^2 - 1}}, \quad 2U'_{n-1}(\tau) = n \frac{\gamma^n_+ + \gamma^n_-}{\tau^2 - 1} - \frac{\tau(\gamma^n_+ - \gamma^n_-)}{(\tau^2 - 1)^{3/2}}.
$$

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

$(3.23)$

$$
2\zeta = \frac{\alpha + \beta}{\alpha\beta} + \frac{2}{\omega \sqrt{\tau^2 - 1}} \frac{n}{\gamma^n_+ + \gamma^n_-} + \frac{2}{\omega \tau^2 - 1} \frac{\tau}{n}
$$

$$
= \frac{2}{\omega \sqrt{\tau^2 - 1}} \frac{n}{\gamma^n_+ + \gamma^n_-} \frac{\gamma^n_+ - \gamma^n_-}{\gamma^n_+ - \gamma^n_-}.
$$

Equation (3.21) and (3.24) imply

$(3.25)$

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

Because $\tau = -((\kappa + 1)/(\kappa - 1),$ $\gamma^{k-1}_+ = (-1)^{k-1} \Delta^{k-1}_\kappa,$ $\gamma^{k-1}_- = (-1)^{k-1} \Delta^{(k-1)}_\kappa,$ and therefore

$(3.26)$

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

For $k = N \equiv n + 1,$ the right-hand side of (3.26) is $2(\Delta^{n}_\kappa + \Delta^{-n}_\kappa)^{-1},$ as was shown by Meinardus [18]. For any other $k,$ we have

$(3.27)$

$$
\min_{|u_{(1)}| = 1} \left\| \text{diag}(g)V_{\kappa,N}^T u \right\|_2 \left\| g \right\|_2 = \rho_{k-1} \times 2 \left[ \Delta^{k-1}_\kappa + \Delta_{\kappa}^{(k-1)} \right]^{-1}.
$$
where
\[
p_{k-1} \overset{\text{def}}{=} \frac{\text{Right-hand side of (3.26)}}{2 \left[ \Delta_{2(k-1)} + \Delta_{(k-1)}^{n} \right]^{-1}}
\]
\[
= \left[ \frac{\left( \Delta_{k-1}^{n} + \Delta_{n}^{(k-1)} \right)^{2}}{4} - \frac{\left( \Delta_{k-1}^{n} - \Delta_{n}^{(k-1)} \right)}{4} \left( \Delta_{n}^{2(k-1)} - \Delta_{n}^{2(1-k)} \right) \right]^{1/2}
\]
\[
= \left[ \frac{1}{2} \left( \Delta_{k-1}^{(k-1)} + 1 \right) \left( \Delta_{n}^{2(k-1)} + 1 \right) \right]^{1/2}
\]
which yields (2.11).

\[\□\]

4. Proof of Theorem 2.2

We first prove two general lemmas for Vandermonde matrix \(V_{N} \equiv V_{N,N}\) as defined in (3.2) with arbitrary, possibly complex, nodes \(\alpha_{j}\).

**Lemma 4.1.** Assume one or more of 1) there are fewer than \(n\) distinct \(\alpha_{j}\), 2) some \(\alpha_{j} = 0\), and 3) some \(g_{(j)} = 0\) occur. Then
\[
\min_{|u_{(1)}| = 1} \| \text{diag}(g)V_{N}^{T}u \|_{2} = \begin{cases} 0, & \text{if all } \alpha_{j} \neq 0; \\ \frac{1}{\sqrt{\sum_{\alpha_{j}=0} |g_{(j)}|^2}}, & \text{otherwise}. \end{cases}
\]

**Proof.** If all \(\alpha_{j} \neq 0\), only Case 1) and 3) are possible. Let \(\ell\) be the number of distinct \(\alpha_{j}\)’s, exclude those corresponding to \(g_{(j)} = 0\). Then \(\ell < n\). By permuting the rows of \(\text{diag}(g)V_{N}^{T}\), we may assume that \(\alpha_{1}, \alpha_{2}, \ldots, \alpha_{\ell}\) are distinct and for \(\alpha_{j} (j > \ell)\) either it is equal to some \(\alpha_{i} (i \leq \ell)\) or corresponding \(g_{(j)} = 0\). Set \(v \in \mathbb{C}^{N}\) whose \(v_{(j)}\) is the coefficient of \(z^{j-1}\) in the polynomial \(\phi(z) = \prod_{j=1}^{n}(z - \alpha_{j})\). Then \(v_{(1)} = \prod_{j=1}^{n}(-\alpha_{j}) \neq 0\), and
\[
\min_{|u_{(1)}| = 1} \| \text{diag}(g)V_{N}^{T}u \|_{2} \leq \| \text{diag}(g)V_{n}^{T}(v/v_{(1)}) \|_{2} = 0,
\]
as expected.

If some \(\alpha_{j} = 0\), then 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. \[\□\]

The next lemma is essentially [17, Theorems 2.1 and 3.1], but stated differently. The proof below has a slightly different flavor.

**Lemma 4.2 ([17]).** Let \(V_{N} \equiv V_{N,N}\) be as defined in (3.2) with all nodes \(\alpha_{j}\) (possibly complex) distinct, and let \(f(z) = \prod_{j=1}^{N}(z - \alpha_{j})\).
If all \( g_{(j)} \neq 0 \), then
\[
\min_{|u_{(1)}|=1} \frac{\|\operatorname{diag}(g)V_N^T u\|_2}{\|u\|_2} = \left[ \sum_{j=1}^{N} \left( \frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right)^2 |g_{(j)}|^{-2} \sum_{j=1}^{N} |g_{(j)}|^2 \right]^{-1/2}.
\]
\[
\max_{g} \min_{|u_{(1)}|=1} \frac{\|\operatorname{diag}(g)V_N^T u\|_2}{\|u\|_2} = \left[ \sum_{j=1}^{N} \left( \frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right)^2 \right]^{-1/2},
\]
where the maximum is achieved if and only if for some constant \( \mu > 0 \),
\[
|g_{(j)}| = \mu \left[ \frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right]^{1/2} \quad \text{for } 1 \leq j \leq N.
\]

**Proof.** In Lemma 3.5, take \( Z = \operatorname{diag}(g)V_N^T \). The assumptions make this \( Z \) nonsingular. Therefore
\[
\left[ \min_{|u_{(1)}|=1} \frac{\|\operatorname{diag}(g)V_N^T u\|_2}{\|u\|_2} \right]^{-2} = e_1^T (\bar{V}_N \Phi V_N^T)^{-1} e_1 = e_1^T (V_N \Phi V_N^T)^{-1} e_1 = (V_N^{-1} e_1)^T \Phi^{-1} (V_N^{-1} e_1),
\]
where \( \Phi = [\operatorname{diag}(g)]^* \operatorname{diag}(g) \) 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
\[
e_1^T (\bar{V}_N \Phi V_N^T)^{-1} e_1 = \sum_{j=1}^{N} \left( \frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right)^2 |g_{(j)}|^{-2},
\]
\[
\min_{|u_{(1)}|=1} \frac{\|\operatorname{diag}(g)V_N^T u\|_2}{\|u\|_2} = \left[ \sum_{j=1}^{N} \left( \frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right)^2 \right]^{-1/2} \sum_{j=1}^{N} |g_{(j)}|^2 \]
\[
\leq \left[ \sum_{j=1}^{N} \left( \frac{|f(0)|}{|\alpha_j| |f'(\alpha_j)|} \right)^2 \right]^{-1/2},
\]
where it is an equality if and only if \( |g_{(j)}| \) are given by (4.3). \( \square \)

**Remark 4.1.** This lemma closely relates to a result of Greenbaum [9, (2.2) and Theorem 1] which in our notation essentially proved that if all nodes \( \alpha_j > 0 \), there exist \( k \) of \( \alpha_j \)'s: \( \alpha_{j_1}, \ldots, \alpha_{j_k} \) such that
\[
\max_{g} \min_{|u_{(1)}|=1} \frac{\|\operatorname{diag}(g)V_k^T u\|_2}{\|u\|_2} = \max_{h} \min_{|u_{(1)}|=1} \frac{\|\operatorname{diag}(h)V_k^T u\|_2}{\|u\|_2}.
\]
where \( V_k \) is the \( k \times k \) Vandermonde matrix with nodes \( \alpha_j, \) \( 1 \leq i \leq k \). 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 [9, 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 \( \alpha_j \).

**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 \( \tau_j \) as in (2.7) with any given \( \omega \) and \( \tau \). Suppose Vandermonde matrix \( V_N \) has nodes \( \alpha_j = \tau_j \) for \( 0 \leq j \leq n \).

(1) If all \( g_j \not\equiv 0 \), then

\[
\begin{align*}
(4.5) \quad \min_{|u_{(i)}| = 1} \frac{\| \text{diag}(g) V_N^T u \|_2}{\| g \|_2} &= \frac{n \omega}{\tau_0 \tau_n, \tau_{n-1}(\tau)} \left[ \sum_{j=1}^{N} |g_j|^2 \right] \times \\
& \left( \frac{1}{(2\tau_0 \tau_n)} |g(1)|^{-2} + \sum_{j=2}^{N-1} \frac{1}{(\tau_j)} |g(j+1)|^{-2} + \frac{1}{(2\tau_n \tau_n)} |g(N)|^{-2} \right)^{-1/2},
\end{align*}
\]

where \( U_{n-1}(t) \) is the \((n-1)\)th Chebyshev polynomial of the second kind as in (3.22).

(2)

\[
\begin{align*}
(4.6) \quad \max_{g} \min_{|u_{(i)}| = 1} \frac{\| \text{diag}(g) V_N^T u \|_2}{\| g \|_2} &= \frac{\frac{n \omega}{|\tau_0 \tau_n, \tau_{n-1}(\tau)|}}{\sum_{j=1}^{n-1} \frac{1}{|\tau_j|} + \frac{1}{(2\tau_n \tau_n)}},
\end{align*}
\]

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

\[
(4.7) \quad |g(j+1)| = \begin{cases} 
\mu \sqrt{1/|\tau_j|}, & \text{for } j \in \{0, n\}, \\
\mu \sqrt{2/|\tau_n|}, & \text{for } 1 \leq j \leq n-1.
\end{cases}
\]

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

\[
f(z) = \eta(z - \tau_0)(z - \tau_n)U_{n-1}(z/\omega + \tau),
\]

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

\[
\begin{align*}
f(0) &= \eta \tau_0 \tau_n U_{n-1}(\tau), \\
f'(\tau_0) &= -\eta (\tau_0 - \tau_n) U_{n-1}(1) \\
&= -\eta (\tau_0 - \tau_n) n \\
&= -\eta 2n\omega, \\
f'(\tau_n) &= -\eta (\tau_n - \tau_0) U_{n-1}(-1) \\
&= (-1)^n \eta (\tau_n - \tau_0) n \\
&= (-1)^n \eta 2n\omega,
\end{align*}
\]
and for $1 \leq j \leq n - 1$

$$f'(x_{jn}^n) = \eta(x_{jn}^n - x_{0n}^n)(x_{jn}^n - x_{nn}^n)U_{n-1}(\tau_{jn})/\omega$$
$$= \eta(x_{jn}^n - x_{0n}^n)(x_{jn}^n - x_{nn}^n)n/\omega(1 - \tau_{jn}^2)$$
$$= -\eta n\omega.$$

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

\[\square\]

**Remark 4.2.** As a corollary to (4.6) and the error bound in (1.4), we deduce that the right-hand side of (4.6) is equal to $|T_n(\tau)| = 2[\Delta_n^k + \Delta_n^{-k}]^{-1}$.

**Proof of Theorem 2.2.** Item 1 is always true for any given positive definite system $Ax = b$. In fact let $A = QAQ^*$ be its eigendecomposition, where $Q$ is unitary, and $A$ as in the theorem since $A$ is positive definite. Set $\overline{g} = \Lambda^{-1/2}Q^*b$. Define $g = (\overline{g}(1), \overline{g}(2), \ldots, \overline{g}(N))^T \in \mathbb{R}^N$. Then $\overline{g} = Dg$ for some diagonal $D$ with $|D(j, j)| = 1$. Finally $A = QAQ^*$ and $b = QA^{1/2}g$ with $Q = \overline{Q}$ still unitary.

Next we notice that

\[(4.8) \| r_k \|_{A^{-1}} = \min_{x \in K_k} \| b - Ax \|_{A^{-1}} \leq \min_{p_k(0) = 1} \| p_k(A)g \|_2 \leq \min_{p_k(0) = 1} \left[ \sum_{j=1}^N |p_k(\lambda_j)|^2 g^2(j) \right]^{1/2}, \]

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 error bound in (1.4) gives

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

contradicting (2.12). This proves Item 2.

For Item 3, we first claim that $\lambda_j$ for which $g(j) > 0$ is in $\{x_{jk}^n, 0 \leq j \leq k\}$. Otherwise if there was a $j_0$ such that $g(j_0) > 0$ and $\lambda_{j_0} \notin \{x_{jk}^n, 0 \leq j \leq k\}$, then $|T_k(\lambda_{j_0}/\omega + \tau)| < 1$, where $\omega$ and $\tau$ 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

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

contradicting (2.12). 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 $\{\lambda_j : g(j) > 0\} \supset \{x_{jk}^n, 0 \leq j \leq k\}$. 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. Kaniel [13] gave a different proof of this fact.
Define \( \hat{g} \in \mathbb{R}^{k+1} \) by \( \hat{g}_{\ell+1} = \| g_J \|_2 \). (4.8) gives

\[
\| r_k \|_2^{-1} = \min_{p_k(0)=1} \left\| \sum_{j=0}^{k} [p_k(\tau_{jk})^2 \hat{g}_{j+1}] \right\|_2 = \min_{\| u \|_1 = 1} \| \text{diag}(\hat{g}) V_{k+1}^T u \|_2,
\]

where \( V_{k+1} \equiv V_{k+1,k+1} \) is the \((k + 1) \times (k + 1)\) Vandermonde matrix as defined in (3.2) with nodes \( \alpha_{j+1} = \tau_{jk} \) for \( 0 \leq j \leq k \). The condition (2.12) and the error bound in (1.4) imply that for \( \hat{g} \)

\[
\min_{\| u \|_1 = 1} \| \text{diag}(\hat{g}) V_{k+1}^T u \|_2 = \max_{\| u \|_1 = 1} \min_{h} \| \text{diag}(h) V_{k+1}^T u \|_2.
\]

Lemma 4.3 shows \( \hat{g}_{\ell+1} = \| g_J \|_2 \) must take the form of (2.13). \( \square \)

5. CONCLUDING REMARKS

We have found a closed formula for the CG residuals for Meinardus’ examples. These residuals may deviate from the well-known error bounds in (1.4) by a factor no bigger than \( 1/\sqrt{2} \), indicating the error bounds by (1.4) 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 \( V_{k}^T = S_k^T R_k^{-1} \), and
3. the solution to \( \min_{\| u \|_1 = 1} \| Z u \|_2 \).

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 [14, 15] 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 residual achieves the error bound in (1.4). Roughly speaking, as far as CG is concerned, these extreme examples are nothing but one of Meinardus’ examples of order \( k + 1 \). As a consequence, unless \( N = 2 \) there is no positive linear system whose \( k \)th CG residual achieves the error bound in (1.4) for all \( 1 \leq k < N \).

ACKNOWLEDGMENT

The author wish to acknowledge the anonymous referee to draw his attention to Chapter II in [5] by H. Rutishauser.

REFERENCES


\[ \text{http://math.univ-lille1.fr/~bbecker/abstract/Habilitationsschrift_Beckermann.pdf}. \]


Department of Mathematics, University of Kentucky, Lexington, KY 40506.

E-mail address: rcli@ms.uky.edu

URL: http://www.ms.uky.edu/~rcli