University of Kentucky Lexington

Department of Mathematics Technical Report 06-01

A Note on Hard Cases For Conjugate Gradient Method[1]

Ren-Cang Li[2]

February 2005
revised October 2005

ABSTRACT

The Conjugate Gradient (CG) method is often used to solve a positive definite linear system $Ax = b$. This paper analyzes two hard cases for CG or any Krylov subspace type methods by either analytically finding the residual formulas or tightly bound the residuals from above and below, in contrast to existing results which only bound residuals from above. The analysis is based on a general framework to estimate CG and MINRES residuals for certain linear systems, and the framework may potentially be useful elsewhere.

[1]This report is available on the web at `http://www.ms.uky.edu/~math/MAreport/`.

[2]Department of Mathematics, University of Kentucky, Lexington, KY 40506 (`rcli@ms.uky.edu`.) 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.

A Note on Hard Cases For Conjugate Gradient Method

Ren-Cang Li *

February 2005
revised October 2005

Abstract

The Conjugate Gradient (CG) method is often used to solve a positive definite linear system $Ax = b$. This paper analyzes two hard cases for CG or any Krylov subspace type methods by either analytically finding the residual formulas or tightly bound the residuals from above and below, in contrast to existing results which only bound residuals from above. The analysis is based on a general framework to estimate CG and MINRES residuals for certain linear systems, and the framework may potentially be useful elsewhere.

1 Introduction

In 1960, Frank [7] tested Richardson's method on an $n \times n$ tridiagonal positive definite linear system $Ax = b$ and experienced very slow convergence. Both Richardson's method and the conjugate gradient method (CG) seek approximations within so-called Krylov subspaces, and satisfy certain optimal properties at each iteration step. What causes the slow convergence? In part, there is no secret about this because intuitively, it has become part of the folklore that CG would have a hard time to solve linear system $Ax = b$ for which A has an eigenvalue distribution similar to that of the zeros of a Chebyshev polynomial (possible after a linear transformation). This is precisely the case for Frank's example. In addition, A in Frank's example has a condition number (the largest eigenvalue over smallest eigenvalue) growing proportional to n^2. The two combined contribute to extreme difficulty for CG to solve the system. Indeed Golub and O'Leary [9] called it "*the hardest case for conjugate gradients*". But can we quantify this slow convergence phenomenon?

There are well-known error bounds due to Meinardus [21] for CG approximations. Frank [7] also presented an error analysis including error bounds for Richardson's method, too. These bounds are in terms of A's condition number, and go to zero very slowly as iteration progresses for the example, as they should. Nevertheless they are upper bounds and, despite the folklore, it would be still of interest to be able to compute exactly or bound tightly the actual residuals from above and below.

In this paper, we first establish a general framework that can be used to compute exactly and/or bound tightly CG and MINRES (the minimal residual method) residuals for certain

*Department of Mathematics, University of Kentucky, Lexington, KY 40506 (rcli@ms.uky.edu.) 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.

linear systems. The key technical detail is very much along the lines of [16, 18], where examples are constructed to show the existing error bound of Meinardus [21] for CG and those of Kaniel-Saad [14, 24] for symmetric Lanczos method are sharp[1] in general, modulo a modest constant factor. It is worth noting that superlinear convergence, much faster than what the existing error bounds suggest, is often observed for both methods [3, 4, 12, 25, 26].

We then apply the framework to analyze the behaviors of CG on Frank's example and a model problem (see, e.g., [6, p.267] and [5]). Exact residual expressions and tight lower and upper bounds are obtained. They shows both problems are indeed difficult, much as expected. The model problem has been widely studied already, mostly as a teaching example to illustrate behaviors of various iterative methods for linear systems. Naiman, Babuška, and Elman [22] computed explicitly CG residuals for 3 special right-hand sides, while [4] Beckermann and Kuijlaars investigated the superlinear convergence behavior of CG when the right-hand sides are from discretizing certain very smooth functions. Our treatment to the model problem appears to be less complicated, with the help of the framework.

The rest of this paper is organized as follows. Section 2 establishes a general framework to analyze the convergence of CG on certain linear systems. The framework is then applied to Frank's example [7] in Section 3 and to the model problem [6] in Section 4. In both examples, exact residuals or tight bounds are obtained. Finally in Section 5 we present our concluding remarks.

Notation. Throughout this paper, $\mathbb{K}^{n\times m}$ is the set of all $n\times m$ matrices with entries in $\mathbb{K}$, where $\mathbb{K}$ is $\mathbb{C}$ (the set of complex numbers) or $\mathbb{R}$ (the set of real numbers), $\mathbb{K}^n=\mathbb{K}^{n\times 1}$, and $\mathbb{K}=\mathbb{K}^1$. 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 jth column.

The inequality $X\le Y$ is for two Hermitian matrices, meaning that $Y-X$ is positive semidefinite. The superscript "$\cdot^*$" takes conjugate transpose while "$\cdot^T$" takes transpose only.

We shall also adopt MATLAB-like convention to access the entries of vectors and matrices. The set of integers from i to j inclusive is $i:j$. For vector u and matrix X, $u_{(j)}$ is u's jth entry, $X_{(i,j)}$ is X's (i,j)th entry, $\mathsf{diag}(u)$ is the diagonal matrix with $(\mathsf{diag}(u))_{(j,j)}=u_{(j)}$, and $\mathsf{diag}(X)$ extracts the diagonal of X; X's submatrices $X_{(k:\ell,i:j)}$, $X_{(k:\ell,:)}$, and $X_{(:,i:j)}$ consists 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 A framework to estimate CG and MINRES residuals for certain linear systems

The kth Krylov subspace $\mathcal{K}_k\equiv\mathcal{K}_k(A,b)$ of A on b is defined as

$$\mathcal{K}_k\equiv\mathcal{K}_k(A,b)\stackrel{\text{def}}{=}\mathsf{span}\{b,Ab,\ldots,A^{k-1}b\}. \tag{2.1}$$

[1]It is well-known that Meinardus' bound for CG *is* sharp (see, e.g., [2, 11, 14, 21]), but it is only claimed *locally* in the sense that for each iteration step k there is a linear system $Ax=b$ (depending on k of course) on which the kth CG residual attains the bound. It turns out that for such $Ax=b$, CG computes the exact solution in the very next iteration [14]! The contribution in [16] is that there is an $Ax=b$ on which the CG residuals are comparable to Meinardus' bound for all iteration steps.

We are interested in the best approximate solution in $\mathcal{K}_k$ to $Ax = b$. For CG, A is positive definite, and the computed solution achieves [6, p.306]

$$\min_{x\in\mathcal{K}_k} \frac{\|A^{-1}b - x\|_A}{\|A^{-1}b\|_A} \equiv \min_{x\in\mathcal{K}_k} \frac{\|b - Ax\|_{A^{-1}}}{\|r_0\|_{A^{-1}}}, \tag{2.2}$$

where the M-vector norm $\|z\|_M \stackrel{\text{def}}{=} \sqrt{z^*Mz}$, $r_0 = b - Ax_0$ with $x_0 = 0$, the initial approximation. For the minimal residual method (MINRES), A may be indefinite or just normal, and the computed solution achieves [23]

$$\min_{x\in\mathcal{K}_k} \frac{\|b - Ax\|_2}{\|r_0\|_2}, \tag{2.3}$$

where the ℓ_2 vector norm $\|z\|_2 = \sqrt{z^*z}$. Let the eigen-decomposition of A be

$$A = U\Lambda U^*,\ U^*U = I_n,\ \Lambda = \mathsf{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \tag{2.4}$$

where all $\lambda_j > 0$ for CG and some λ_j may be complex for MINRES. It has been verified that both (2.2) and (2.3) can be restated as [13, 16, 20]

$$\min_{|u_{(1)}|=1} \frac{\|\mathsf{diag}(g)\, V_{k+1,n}^T u\|_2}{\|g\|_2}, \tag{2.5}$$

where $g = \Lambda^{-1/2}U^*b$ for CG, and $g = U^*b$ for MINRES, and

$$V_{k+1,n} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^k & \lambda_2^k & \cdots & \lambda_n^k \end{pmatrix}. \tag{2.6}$$

The quantity (2.5) points to a new direction to analyze the convergence behavior of CG or MINRES on certain linear systems. Frank's example and the model problem are two of them.

Now we are ready to explain our general framework for solving (2.5). The key idea is similar to that in [16], i.e., seeking a decomposition, reminiscent of QR, for $V_{k+1,n}^T$ but what presented here is in a more general term. The framework is built upon

- a family of polynomials $p_m(t)$ of degree m in t for $m = 0, 1, 2, ...$,
- two parameters $\omega \neq 0$ and τ (complex or real), and
- $\Omega_n = \mathsf{diag}(\omega_1, \omega_2, \dots, \omega_n)$ with all $\omega_j > 0$.

(2.7)

For the family of polynomials, we often pick Chebyshev polynomials of the 1st kind:

$$\begin{aligned} T_m(t) &= \cos(m \arccos t) && \text{for } |t| \le 1, && (2.8) \\ &= \frac{1}{2}\left(t + \sqrt{t^2-1}\right)^m + \frac{1}{2}\left(t - \sqrt{t^2-1}\right)^m && \text{for } |t| \ge 1, && (2.9) \end{aligned}$$

or more generally any orthogonal polynomials, as in our later applications; usually $\Omega_n = I_n$ or take $\{\omega_j\}_{j=1}^n$ to be weights in the associated Gaussian quadrature rule if p_m is an orthogonal polynomial; the selections of $\omega \neq 0$ and τ can be tricky and are obviously problem-dependent.

First we define the mth *Translated Polynomial* in x of degree m as[2]

$$\begin{aligned} p_m(x;\omega,\tau) &\stackrel{\text{def}}{=} p_m(x/\omega+\tau) && (2.10)\\ &= a_{mm}x^m + a_{m-1\,m}x^{m-1} + \cdots + a_{1m}x + a_{0m}, && (2.11)\end{aligned}$$

where $a_{jm} \equiv a_{jm}(\omega,\tau)$ are functions of ω and τ, and upper triangular $R_m \in \mathbb{C}^{m\times m}$, a matrix-valued function in ω and τ, too, as

$$R_m \equiv R_m(\omega,\tau) \stackrel{\text{def}}{=} \begin{pmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0\,m-1} \\ & a_{11} & a_{12} & \cdots & a_{1\,m-1} \\ & & a_{22} & \cdots & a_{2\,m-1} \\ & & & \ddots & \vdots \\ & & & & a_{m-1\,m-1} \end{pmatrix}, \tag{2.12}$$

i.e., the jth column consists of the coefficients of $p_{j-1}(x;\omega,\tau)$. Denote $V_n = V_{n,n}$ and set

$$\boldsymbol{P}_n \stackrel{\text{def}}{=} \begin{pmatrix} p_0(t_1) & p_0(t_2) & \cdots & p_0(t_n) \\ p_1(t_1) & p_2(t_2) & \cdots & p_1(t_n) \\ \vdots & \vdots & & \vdots \\ p_{n-1}(t_1) & p_{n-1}(t_2) & \cdots & p_{n-1}(t_n) \end{pmatrix}, \tag{2.13}$$

where

$$t_j = \lambda_j/\omega + \tau \quad \text{for } 1 \le j \le n. \tag{2.14}$$

Then

$$V_n^T R_n = \boldsymbol{P}_n^T, \tag{2.15}$$

Equation (2.15) yields $V_n^T = \boldsymbol{P}_n^T R_n^{-1}$. Extracting the first $k+1$ columns from both sides of $V_n^T = \boldsymbol{P}_n^T R_n^{-1}$ yields

$$V_{k+1,n}^T = \boldsymbol{P}_{k+1,n}^T R_{k+1}^{-1}, \tag{2.16}$$

where $\boldsymbol{P}_{k+1,n} = \boldsymbol{P}_{(1:k+1,:)}$. Next set, with matrix Ω_n in (2.7),

$$\Upsilon \stackrel{\text{def}}{=} \boldsymbol{P}_n \Omega_n \boldsymbol{P}_n^*. \tag{2.17}$$

A key step that makes our framework work is the following assumption.

$$\boxed{\begin{array}{l} \text{there are } 0 < \gamma_{k+1} \le \Gamma_{k+1} \text{ and } D_{k+1} = \mathsf{diag}(d_1, d_2, \ldots) \text{ with all } d_j > 0 \\ \text{such that} \\ \qquad\qquad \gamma_{k+1} D_{k+1}^{-1} \le (\Upsilon)_{(1:k+1,1:k+1)} \le \Gamma_{k+1} D_{k+1}^{-1}. \end{array}} \tag{2.18}$$

[2]The symbol x is overloaded. It is a vector in $Ax = b$ and a scalar variable in $p_m(x;\omega,\tau)$. This overuse of x will unlikely cause any confusion, however.

Such γ_{k+1}, Γ_{k+1}, and D_{k+1} always exist if $(\Upsilon)_{(1:k+1,1:k+1)}$ is nonsingular. For example, $D_{k+1} = I_{k+1}$, γ_{k+1} and Γ_{k+1} are the smallest and largest eigenvalues of $(\Upsilon)_{(1:k+1,1:k+1)}$. In using our framework later, they are often estimated. It can be proved that[3] [13, 16, 19]

$$\min_{|u_{(1)}|=1} \|\Omega_n^{1/2} V_{k+1,n}^T u\|_2 = \left[e_1^T (\bar{V}_{k+1,n} \Omega_n V_{k+1,n}^T)^{-1} e_1\right]^{-1/2}, \tag{2.19}$$

where $\bar{V}_{k+1,n}$ is the complex conjugate of $V_{k+1,n}$. By (2.16), we have

$$\begin{aligned} \bar{V}_{k+1,n} \Omega_n V_{k+1,n}^T &= R_{k+1}^{-*} \left[\boldsymbol{P}_{k+1,n}^T\right]^* \Omega_n \boldsymbol{P}_{k+1,n}^T R_{k+1}^{-1} \\ &= R_{k+1}^{-*} \left(\bar{\boldsymbol{P}}_n \Omega_n \boldsymbol{P}_n^T\right)_{(1:k+1,1:k+1)} R_{k+1}^{-1} \\ &= R_{k+1}^{-*} \left(\bar{\Upsilon}\right)_{(1:k+1,1:k+1)} R_{k+1}^{-1}, &(2.20) \\ (\bar{V}_{k+1,n} \Omega_n V_{k+1,n}^T)^{-1} &= R_{k+1} \left[(\bar{\Upsilon})_{(1:k+1,1:k+1)}\right]^{-1} R_{k+1}^*. &(2.21) \end{aligned}$$

Consequently, by (2.18),

$$\frac{1}{\Gamma_{k+1}} e_1^T R_{k+1} D_{k+1} R_{k+1}^* e_1 \le e_1^T (\bar{V}_{k+1,n} \Omega_n V_{k+1,n}^T)^{-1} e_1 \le \frac{1}{\gamma_{k+1}} e_1^T R_{k+1} D_{k+1} R_{k+1}^* e_1$$

which implies

$$\frac{1}{\sqrt{\sum_{j=1}^n \omega_j}} \sqrt{\frac{\gamma_{k+1}}{\Phi_{k+1}}} \le \min_{|u_{(1)}|=1} \frac{\|\Omega_n^{1/2} V_{k+1,n}^T u\|_2}{\sqrt{\sum_{j=1}^n \omega_j}} \le \frac{1}{\sqrt{\sum_{j=1}^n \omega_j}} \sqrt{\frac{\Gamma_{k+1}}{\Phi_{k+1}}}, \tag{2.22}$$

where

$$\Phi_{k+1} = e_1^T R_{k+1} D_{k+1} R_{k+1}^* e_1 = \sum_{j=0}^k d_j |a_{0j}|^2 = \sum_{j=0}^k d_j |p_j(\tau)|^2. \tag{2.23}$$

Occasionally it may also be possible to calculate $\min_{|u_{(1)}|=1} \|\Omega_n^{1/2} V_{k+1,n}^T u\|_2$ exactly as it happens to the two examples we will be dealing with.

We summarize our findings into two mathematically equivalent theorems, in terms of $V_{k+1,n}$ or the residual for $Ax = b$, respectively.

Theorem 2.1 *Given $V_{k+1,n}$ defined as in* (2.6)*, and given, as in* (2.7)*, the polynomial family $\{p_m\}_{m=0}^\infty$ and two prescribed numbers $\omega \neq 0$ and τ and Ω_n. Let t_j be as in* (2.14)*, $\boldsymbol{P}_n$ as in* (2.13)*, and Υ as in* (2.17)*. Then*

$$\min_{|u_{(1)}|=1} \frac{\|\Omega_n^{1/2} V_{k+1,n}^T u\|_2}{\sqrt{\sum_{j=1}^n \omega_j}} = \frac{1}{\sqrt{\sum_{j=1}^n \omega_j}} \frac{1}{\sqrt{y^* \left[(\Upsilon)_{(1:k+1,1:k+1)}\right]^{-1} y}}, \tag{2.24}$$

where $y = R_{k+1}^* e_1 = (p_0(\tau), p_1(\tau), \ldots, p_k(\tau))^*$. If (2.18) *holds, then we have* (2.22)*.*

Remark **2.1** Lower and upper bounds on (2.5) for any general $g \in \mathbb{C}^n$ are immediately available once (2.5) for $g = \mathsf{diag}(\Omega_n^{1/2})$ is known or bounded, since

$$\|\mathsf{diag}(g)\, V_{k+1,n}^T u\|_2 = \|\mathsf{diag}(g) \Omega_n^{-1/2}\, \Omega_n^{1/2} V_{k+1,n}^T u\|_2$$

[3] Equation (2.19), though not explicitly stated this way in [13, 19], is implicitly implied there.

which yields

$$\left(\min_j \frac{|g_{(j)}|}{\sqrt{\omega_j}}\right) \|\Omega_n^{1/2} V_{k+1,n}^T u\|_2 \le \|\mathsf{diag}(g)\, V_{k+1,n}^T u\|_2 \le \left(\max_j \frac{|g_{(j)}|}{\sqrt{\omega_j}}\right) \|\Omega_n^{1/2} V_{k+1,n}^T u\|_2. \quad (2.25)$$

For this reason, in the rest of this paper, we shall simply consider (2.5) for $g = \mathsf{diag}(\Omega_n^{1/2})$.

Theorem 2.2 *Let $A \in \mathbb{C}^{n\times n}$ be normal and have eigen-decomposition* (2.4)*, $b \in \mathbb{C}^n$, and suppose* (2.7)*. Define t_j as in* (2.14)*, $\boldsymbol{P}_n$ as in* (2.13)*, and Υ as in* (2.17)*. For normal A, $q = 2$ and $g = U^* b$, and for positive definite A, either $q = 2$ and $g = U^* b$ or $q = A^{-1}$ and $g = \Lambda^{-1/2} U^* b$. If $|g_{(j)}| = \omega_j^{1/2}$ for $1 \le j \le n$, then*

$$\min_{x\in\mathcal{K}_k} \frac{\|b - Ax\|_q}{\|r_0\|_q} = \frac{1}{\sqrt{\sum_{j=1}^n \omega_j}} \frac{1}{\sqrt{y^* \left[(\Upsilon)_{(1:k+1,1:k+1)}\right]^{-1} y}}, \quad (2.26)$$

where $y = R_{k+1}^* e_1 = (p_0(\tau), p_1(\tau), \ldots, p_k(\tau))^*$. If (2.18) *holds, then*

$$\left(\min_j \frac{|g_{(j)}|}{\sqrt{\omega_j}}\right) \frac{1}{\|g\|_2} \sqrt{\frac{\gamma_{k+1}}{\Phi_{k+1}}} \le \min_{x\in\mathcal{K}_k} \frac{\|b - Ax\|_q}{\|r_0\|_q} \le \left(\max_j \frac{|g_{(j)}|}{\sqrt{\omega_j}}\right) \frac{1}{\|g\|_2} \sqrt{\frac{\Gamma_{k+1}}{\Phi_{k+1}}}. \quad (2.27)$$

Remark **2.2** When t_j are the zeros of the nth polynomial from an orthogonal polynomial family, one should take ω_j to be the weights of the associated Gaussian quadrature rule, and then $\boldsymbol{P}_n \Omega_n \boldsymbol{P}_n^*$ is diagonal since the Gaussian quadrature rule is exact for polynomial of degree no higher than $2n-1$. Then $\min_{|u_{(1)}|=1} \|\Omega_n^{1/2} V_{k+1,n}^T u\|_2$ has a close formula solution.

Remark **2.3** Besides yielding bounds on residuals for $Ax = b$, (2.21) may have other independent interests. For example, it relates $\mathrm{trace}(\bar{V}_{k+1,n}\Omega_n V_{k+1,n}^T)^{-1}$ to the coefficients of $p_j(x;\omega,\tau)$ by

$$\Gamma_{k+1}^{-1}\|R_{k+1}\|_{\mathrm{F}}^2 \le \mathrm{trace}\left((\bar{V}_{k+1,n}\Omega_n V_{k+1,n}^T)^{-1}\right) \le \gamma_{k+1}^{-1}\|R_{k+1}\|_{\mathrm{F}}^2. \quad (2.28)$$

It is natural to define the condition number in Ω_n-Frobenius norm of $V_{k+1,n}$ as

$$\kappa_{\mathrm{F},\Omega_n}(V_{k+1,n}) \stackrel{\mathrm{def}}{=} \|\Omega_n^{1/2} V_{k+1,n}\|_{\mathrm{F}} \sqrt{\mathrm{trace}\left((\bar{V}_{k+1,n}\Omega_n V_{k+1,n}^T)^{-1}\right)}.$$

$\|\Omega_n^{1/2} V_{k+1,n}\|_{\mathrm{F}}$ can be easily computed (estimated). If $\|R_{k+1}\|_{\mathrm{F}}$ can be estimated, too, bounds on $\kappa_{\mathrm{F},\Omega_n^{1/2}}(V_{k+1,n})$ can be obtained. Indeed for $p_j = T_j$, if τ is real and $\tau = 0$ or $|\tau| \ge 1$, tight bounds on $\|R_{k+1}\|_{\mathrm{F}}$ are indeed available [15] (see also [17]).

For future reference, Meinardus' bound [21], restated in terms of $V_{k+1,n}$, is: *Assume all $\lambda_j > 0$ and set $\kappa = \max_j \lambda_j / \min_j \lambda_j$. Then*

$$\min_{|u_{(1)}|=1} \frac{\|\mathsf{diag}(g)\, V_{k+1,n}^T u\|_2}{\|g\|_2} \le 2\left[\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^{-k}\right]^{-1}. \quad (2.29)$$

The right-hand side is in fact $|T_k(t)|^{-1}$ at $t = (\kappa+1)/(\kappa-1)$.

3 Frank's example

In Frank's example [7],

$$A = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \in \mathbb{R}^{n\times n}. \tag{3.1}$$

A admits an analytical eigen-decomposition (2.4). Frank [7] credited this close expression

$$\theta_j = \frac{2j-1}{2n+1}\pi, \quad \lambda_j = 2 - 2\cos\theta_j \quad \text{for } 1 \le j \le n \tag{3.2}$$

to Young [27] who, presumably, found closed formulas for the associated eigenvectors, too. But we do not have access to Young [27]. In any case, it can be computed that the entries of the jth eigenvector are

$$(U)_{(i,j)} = \frac{2}{\sqrt{2n+1}} \cos\left(\pi\frac{(2i-1)(2j-1)}{2(2n+1)}\right) \quad \text{for } 1 \le i \le n.$$

For this example, we shall take $p_m = T_m$, $\omega = -2$, $\tau = 1$, and $\Omega_n = I_n$ in the framework in Section 2.

Lemma 3.1 *For $\theta_j = \frac{2j-1}{2n+1}\pi$ and integer ℓ,*

$$\sum_{k=1}^{n} \cos \ell\theta_k = \begin{cases} (-1)^m n, & \text{if } \ell = m(2n+1) \text{ for some integer } m, \\ 1/2, & \text{if } \ell \text{ is odd, but } \ell \ne m(2n+1) \text{ for any integer } m, \\ -1/2, & \text{if } \ell \text{ is even, but } \ell \ne m(2n+1) \text{ for any integer } m. \end{cases} \tag{3.3}$$

Proof: If $\ell = m(2n+1)$ for some integer m, then $\ell\theta_k = m(2k-1)\pi$ and thus $\cos \ell\theta_k = (-1)^m$. Assume that $\ell \ne m(2n+1)$ for any integer m. Set $\phi = \ell\pi/(2n+1)$. We have [10, p.30]

$$\sum_{k=1}^{n} \cos \ell\theta_k = \sum_{k=1}^{n} \cos(2k-1)\phi = \frac{1}{2}\frac{\sin 2n\phi}{\sin\phi}.$$

Now notice $\sin 2n\phi = \sin(\ell\pi - \phi) = -(-1)^\ell \sin\phi$ to conclude the proof. ∎

Lemma 3.2 *Set $p_m = T_m$, $\omega = -2$, $\tau = 1$, and $\Omega_n = I_n$ in (2.7). Let $\Upsilon = \boldsymbol{P}_n\Omega_n\boldsymbol{P}_n^T$ be defined as in (2.17), and let $w \in \mathbb{R}^n$ whose odd entries are 1 and even entries are -1. Then*

$$\Upsilon = D - \frac{1}{2}ww^T, \quad D = \frac{2n+1}{4}\,\mathsf{diag}(2,1,1,\ldots). \tag{3.4}$$

Allow D and w to have generic sizes but otherwise the same. We still have

$$(\Upsilon)_{(1:k+1,1:k+1)} = D - \frac{1}{2}ww^T = D^{1/2}\left(I_{k+1} - \frac{1}{2}D^{-1/2}ww^TD^{-1/2}\right)D^{1/2}, \tag{3.5}$$

and thus

$$\frac{2(n-k)}{2n+1}D \le (\Upsilon)_{(1:k+1,1:k+1)} \le D. \tag{3.6}$$

Proof: For $0 \le i, j \le n-1$,

$$\begin{aligned}
(\boldsymbol{P}_n \boldsymbol{P}_n^T)_{(i+1,j+1)} &= \sum_{k=1}^{n} (\boldsymbol{P}_n)_{(i+1,k)} (\boldsymbol{P}_n^T)_{(k,j+1)} \\
&= \sum_{k=1}^{n} T_i(t_k) T_j(t_k) \\
&= \sum_{k=1}^{n} \cos i\theta_k \cos j\theta_k \\
&= \frac{1}{2} \sum_{k=1}^{n} \cos(i+j)\theta_k + \frac{1}{2} \sum_{k=1}^{n} \cos(i-j)\theta_k. \qquad (3.7)
\end{aligned}$$

Consider $\ell = i \pm j$ in Lemma 3.1, and $0 \le i,\ j \le n-1$. Since $0 \le i+j \le 2n-2$ and $-n+1 \le i-j \le n-1$, for some integer m

$$i + j = m(2n+1) \quad \Leftrightarrow \quad i = j = 0; \qquad (3.8)$$

$$i - j = m(2n+1) \quad \Leftrightarrow \quad i = j. \qquad (3.9)$$

It can then be verified that

$$(\Upsilon)_{(i+1,j+1)} = \begin{cases} n, & \text{for } i = j = 0, \\ \frac{2n-1}{4}, & \text{for } 0 \ne i = j, \\ -1/2, & \text{for } i \ne j, \text{ both odd or even}, \\ \;\;1/2, & \text{for } i \ne j, \text{ one odd and one even}. \end{cases} \qquad (3.10)$$

This proves (3.4) and (3.5). Next we notice $\|D^{-1/2}w\|_2^2 = \frac{2}{2n+1} + k \times \frac{4}{2n+1} = \frac{2(2k+1)}{2n+1}$ to get

$$I_{k+1} - \frac{1}{2} D^{-1/2} w w^T D^{-1/2} \ge \left(1 - \frac{1}{2}\|D^{-1/2}w\|_2^2\right) I_{k+1} = \frac{2(n-k)}{2n+1} I_{k+1},$$

and thus $(\Upsilon)_{(1:k+1,1:k+1)} \ge \frac{2(n-k)}{2n+1} D$. On the other hand, $(\Upsilon)_{(1:k+1,1:k+1)} \le D$. ∎

Theorem 3.1 *For $V_{k+1,n}$ with nodes $\alpha_j = \lambda_j$, the eigenvalues* (3.2) *of Frank matrix* (3.1),

$$\frac{1}{\sqrt{n}} \sqrt{\frac{n-k}{2k+1}} \le \min_{|u_{(1)}|=1} \frac{\|V_{k+1,n}^T u\|_2}{\sqrt{n}} \le \frac{1}{\sqrt{n}} \sqrt{\frac{2n+1}{2(2k+1)}}, \qquad (3.11)$$

$$\min_{|u_{(1)}|=1} \frac{\|V_{k+1,n}^T u\|_2}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sqrt{\frac{(2n+1)(n-k)}{2(2k+1)(n-k)+1}}. \qquad (3.12)$$

The right-hand side of (3.12) *is very close to the upper bound in* (3.11).

Proof: Set $p_m = T_m$, $\omega = -2$, $\tau = 1$, and $\Omega_n = I_n$ in (2.7). Lemma 3.2 says that (2.18) holds with $\gamma_{k+1} = \frac{2(n-k)}{2n+1}$, $\Gamma_{k+1} = 1$, and $D_{k+1} = D^{-1}$ which, substituted into (2.22) and (2.23), yield (3.11), since

$$\Phi_{k+1} = \sum_{j=0}^{k} d_j |T_j(1)|^2 = \frac{2(2k+1)}{2n+1}.$$

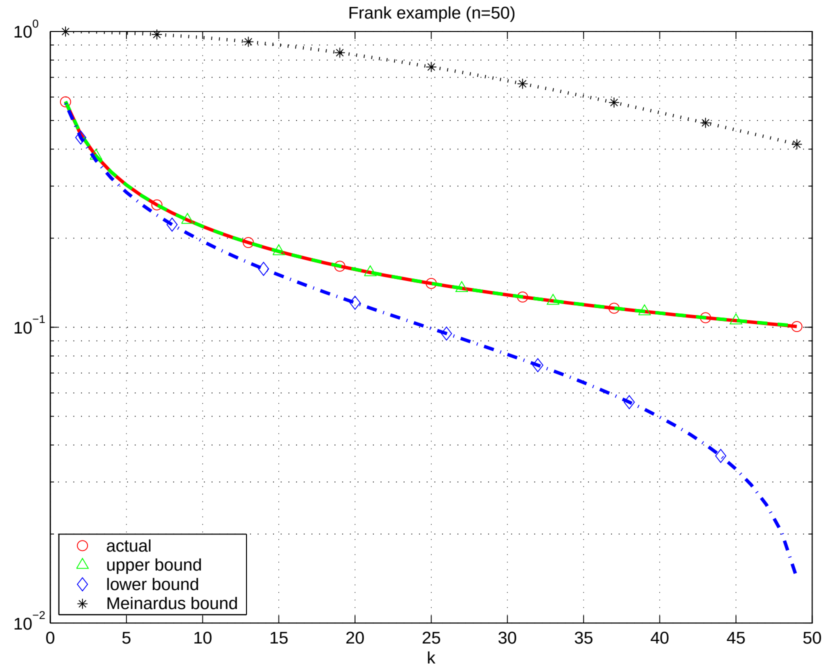

Figure 3.1: Residual $\min_{|u_{(1)}|=1} \frac{\|V_{k+1,n}^T u\|_2}{\sqrt{n}}$ *vs.* its lower and upper bounds for Frank's example: upper bounds nearly the same as actuals.

To prove (3.12), we have, by Sherman-Morrison formula [6, p.95],

$$\begin{aligned} \left[(\Upsilon)_{(1:k+1,1:k+1)}\right]^{-1} &= D^{-1} + \frac{1}{2} D^{-1} w w^T D^{-1} \frac{1}{1 - w^T D^{-1} w/2} \\ &= \frac{4}{2n+1} \left(\mathsf{diag}(2^{-1}, 1, 1, \dots) + \frac{1}{n-k} z z^T \right), \end{aligned}$$

where $z = (2^{-1}, -1, 1, -1, 1, \dots)^T$. Let $y = R_{k+1}^T e_1 = (1, 1, \dots, 1)^T$, and then $y^T z = (-1)^k/2$. By (2.21), we have

$$\begin{aligned} e_1^T (V_{k+1,n} V_{k+1,n}^T)^{-1} e_1 &= y^T \left[(\Upsilon)_{(1:k+1,1:k+1)}\right]^{-1} y \\ &= \frac{4}{2n+1} \left(\frac{1}{2} + k + \frac{1}{n-k} \frac{1}{4} \right) \\ &= \frac{2(2k+1)(n-k)+1}{(2n+1)(n-k)}, \end{aligned}$$

and hence (3.12) because of (2.19). ∎

(3.11) and (3.12) indicate slow convergence of any method that seeks approximations from $\mathcal{K}_k(A, b)$ (if $g = U^* b$ or $\Lambda^{-1/2} U^* b$ does not have extremely tiny entries). Figure 3.1 plots the exact value in (3.12), its lower and upper bounds by (3.11), and the upper bound (2.29) of Meinardus. Notice that the relative residuals can only be reduced by as much as a factor of about 10^{-1} at $k = 49$ in Figure 3.1 (and of course the exact solution at $k = 50$ since then the search space is the entire $\mathbb{C}^n$ – a situation that is not very interesting practically).

We commented at the beginning of this paper that there are actually two reasons contributing to the slow convergence by any method on this A that seeking approximations from $\mathcal{K}_k(A,b)$ – the eigenvalue distribution and the condition number. Let us now examine the two. The zeros of translated Chebyshev polynomial $T_n(x/\omega+\tau) = T_n(1-x/2)$ are $\left\{2-2\cos\frac{2j-1}{2n}\pi\right\}_{j=1}^{n}$ distribute on $(0,4)$ in much the same way as the eigenvalues $\left\{2-2\cos\frac{2j-1}{2n+1}\pi\right\}_{j=1}^{n}$ here in the example for large n. At the same time, the condition number $\kappa(A)=\lambda_n/\lambda_1 \approx 16n^2/\pi^2$ rapidly increases proportionally to n^2.

4 The model problem

Discretizing the following one-dimensional Poisson's equation

$$
\begin{aligned}
&-\frac{d^2v(x)}{dx^2} = f(x), \quad \text{for } 0<x<1,\\
&v(0)=v(1)=0,
\end{aligned}
$$

leads to a linear system $Ax=b$, so-called *the model problem* [6, p.267], with

$$
A = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \in \mathbb{R}^{n\times n}. \tag{4.1}
$$

A admits an analytical eigen-decomposition (2.4), too. In fact [6, 22],

$$
\theta_j = \frac{j}{n+1}\pi, \quad \lambda_j = 2-2\cos\theta_j \quad \text{for } 1\le j\le n \tag{4.2}
$$

and the corresponding jth eigenvector is

$$
(U)_{(i,j)} = \frac{2}{\sqrt{n+1}}\sin\left(\pi\frac{ij}{n+1}\right) \quad \text{for } 1\le i\le n.
$$

For this example, we shall take, in Section 2, either $p_m = T_m$ or U_m (the Chebyshev polynomial of the 2nd kind), $\omega=-2$, $\tau=1$, and either $\Omega_n = I_n$ or $\Omega_n = \mathsf{diag}(\sin\theta_1,\dots,\sin\theta_n)$.

In [22], three close expressions for CG residuals for three special right-hand sides were explicitly computed through clever, albeit complicated, trigonometric function manipulations. In our notation, equivalently, the computations there are for minimization problem (2.5) with

$$
g_{(j)} = 2\sin^{1-\beta}\frac{\theta_j}{2},\ 1\le j\le n, \quad \text{for } \beta=0,1,2. \tag{4.3}
$$

In this section, we shall apply our general framework to solve (2.5) for[4] $|g_{(j)}|=2$ corresponding to $\beta=1$ [22]. This will provide a different approach that appears to be less complicated.

Recall Remark 2.2 for t_j being the zeros of an orthogonal polynomial. It turns out t_j here are the zeros of the nth Chebyshev polynomial $U_n(t)$ of the second kind [1, p.889]. This enables us to produce close formulas for CG residuals for a particular g, other than those in (4.3) considered by [22], in a much simpler way. We shall do so in Theorem 4.2.

[4]Since the constant factor 2 can and will be cancelled, we actually consider equivalently the case $|g_{(j)}|=1$.

Lemma 4.1 *For $\theta_j = \frac{j}{n+1}\pi$ and integer ℓ,*

$$\sum_{k=1}^{n} \cos \ell\theta_k = \begin{cases} n, & \text{if } \ell = 2m(n+1) \text{ for some integer } m, \\ 0, & \text{if } \ell \text{ is odd,} \\ -1, & \text{if } \ell \text{ is even, but } \ell \neq 2m(n+1) \text{ for any integer } m. \end{cases} \tag{4.4}$$

Proof: If $\ell = 2m(n+1)$ for some integer m, then $\ell\theta_k = 2mk\pi$ and thus $\cos \ell\theta_k = 1$. Assume that $\ell \neq 2m(n+1)$ for any integer m. Set $\phi = \ell\pi/(n+1)$. We have [10, p.30]

$$\sum_{k=1}^{n} \cos \ell\theta_k = \sum_{k=1}^{n} \cos k\phi = \cos \frac{n+1}{2}\phi \times \frac{\sin \frac{n\phi}{2}}{\sin \frac{\phi}{2}}.$$

Now notice $\cos \frac{n+1}{2}\phi = \cos \frac{\ell}{2}\pi = 0$ for odd ℓ and $(-1)^{\ell/2}$ for even ℓ, and $\sin \frac{n\phi}{2} = \sin(\frac{\ell}{2}\pi - \phi) = -(-1)^{\ell/2} \sin\phi$ for even ℓ to conclude the proof. ∎

Lemma 4.2 *Set $p_m = T_m$, $\omega = -2$, $\tau = 1$, and $\Omega_n = I_n$ in (2.7). Let $\Upsilon = \boldsymbol{P}_n \Omega_n \boldsymbol{P}_n^T$ be defined as in (2.17), and let e_{odd} and e_{even} be two column vectors defined as: all the odd entries of e_{odd} are ones and the even entries are zeros, and all the odd entries of e_{even} are zeros and the even entries are ones. Then*

$$\Upsilon = D - e_{\mathsf{odd}} e_{\mathsf{odd}}^T - e_{\mathsf{even}} e_{\mathsf{even}}^T, \quad D = \frac{n+1}{2} \mathsf{diag}(2, 1, 1, \ldots). \tag{4.5}$$

Allow D, e_{odd}, and e_{even} to have generic sizes but otherwise the same. We still have

$$(\Upsilon)_{(1:k+1,1:k+1)} = D - e_{\mathsf{odd}} e_{\mathsf{odd}}^T - e_{\mathsf{even}} e_{\mathsf{even}}^T \tag{4.6}$$

$$= D^{1/2} \left[I_{k+1} - \frac{1}{2} D^{-1/2} (e_{\mathsf{odd}} e_{\mathsf{odd}}^T + e_{\mathsf{even}} e_{\mathsf{even}}^T) D^{-1/2} \right] D^{1/2}, \tag{4.7}$$

and thus

$$\frac{n - 2\lfloor k/2 \rfloor}{n+1} D \leq (\Upsilon)_{(1:k+1,1:k+1)} \leq D. \tag{4.8}$$

Proof: For $0 \leq i, j \leq n-1$, (3.7) with θ_j defined here remains valid. Consider now $\ell = i \pm j$, and $0 \leq i,\ j \leq n-1$. Since $0 \leq i+j \leq 2n-2$ and $-n+1 \leq i-j \leq n-1$, for some integer m

$$i + j = 2m(n+1) \quad \Leftrightarrow \quad i = j = 0; \tag{4.9}$$

$$i - j = 2m(n+1) \quad \Leftrightarrow \quad i = j. \tag{4.10}$$

It can then be verified that

$$(\Upsilon)_{(i+1,j+1)} = \begin{cases} n, & \text{for } i = j = 0, \\ \frac{n-1}{2}, & \text{for } 0 \neq i = j, \\ -1, & \text{for } i \neq j, \text{ both odd or even,} \\ 0, & \text{for } i \neq j, \text{ one odd and one even.} \end{cases} \tag{4.11}$$

This gives (4.5) – (4.7). Let $\lfloor k/2 \rfloor$ be the largest integer that is no bigger than $k/2$, and $\lceil k/2 \rceil$ the smallest integer that is no less than $k/2$. Note that $D^{-1/2} e_{\mathsf{odd}}$ and $D^{-1/2} e_{\mathsf{even}}$ are orthogonal, and that

$$\|D^{-1/2} e_{\mathsf{odd}}\|_2^2 = \frac{1}{n+1} + \frac{2}{n+1} \left\lfloor \frac{k}{2} \right\rfloor,$$

$$\|D^{-1/2} e_{\mathsf{even}}\|_2^2 = \frac{2}{n+1} \left\lceil \frac{k}{2} \right\rceil,$$

to get (notice that $\|D^{-1/2}e_{\mathsf{odd}}\|_2^2 \geq \|D^{-1/2}e_{\mathsf{even}}\|_2^2$)

$$I_{k+1} - \frac{1}{2}D^{-1/2}(e_{\mathsf{odd}}e_{\mathsf{odd}}^T + e_{\mathsf{even}}e_{\mathsf{even}}^T)D^{-1/2} \geq \left(1 - \frac{2\lfloor k/2\rfloor + 1}{n+1}\right) I_{k+1} = \frac{n - 2\lfloor k/2\rfloor}{n+1} I_{k+1},$$

and thus $(\Upsilon)_{(1:k+1,1:k+1)} \geq \frac{n-2\lfloor k/2\rfloor}{n+1}D$. On the other hand, $(\Upsilon)_{(1:k+1,1:k+1)} \leq D$. ∎

Theorem 4.1 *For $V_{k+1,n}$ with nodes $\alpha_j = \lambda_j$, the eigenvalues* (4.2) *of model matrix* (4.1),

$$\frac{1}{\sqrt{n}}\sqrt{\frac{n-2\lfloor k/2\rfloor}{2k+1}} \leq \min_{|u_{(1)}|=1} \frac{\|V_{k+1,n}^T u\|_2}{\sqrt{n}} \leq \frac{1}{\sqrt{n}}\sqrt{\frac{n+1}{2k+1}}, \tag{4.12}$$

$$\min_{|u_{(1)}|=1} \frac{\|V_{k+1,n}^T u\|_2}{\sqrt{n}} = \frac{1}{\sqrt{n}}\left[\frac{2k+1}{n+1} + \frac{1}{n+1}\left(\frac{(k+1)^2}{n-k} + \frac{k^2}{n-k+1}\right)\right]^{-1/2}. \tag{4.13}$$

The right-hand side of (4.13) *is very close to the lower bound in* (4.12).

Proof: Set $p_m = T_m$, $\omega = -2$, $\tau = 1$, and $\Omega_n = I_n$ in (2.7). Lemma 4.2 says that (2.18) holds with $\gamma_{k+1} = \frac{n-2\lfloor k/2\rfloor}{n+1}$, $\Gamma_{k+1} = 1$, and $D_{k+1} = D^{-1}$ which, substituted into (2.22) and (2.23), yield (4.12) since $\Phi_{k+1} = \sum_{j=0}^k d_j|T_j(1)|^2 = \frac{2k+1}{n+1}$.

For $k \geq 1$, by Sherman-Morrison-Woodbury formula [6, p.95],

$$\left[(\Upsilon)_{(1:k+1,1:k+1)}\right]^{-1} = D^{-1} + D^{-1}W(I_2 - W^TD^{-1}W)^{-1}W^TD^{-1},$$

where $W = (e_{\mathsf{odd}}\,,\, e_{\mathsf{even}})$. Let $y = R_{k+1}^Te_1 = (1,1,\ldots,1)^T$. We need to evaluate $y^T\left[(\Upsilon)_{(1:k+1,1:k+1)}\right]^{-1}y$. To this end, we notice

$$\begin{aligned}
y^TD^{-1} &= \frac{2}{n+1}(2^{-1},1,1,\ldots),\\
y^TD^{-1}y^T &= \frac{2}{n+1}\left(\frac{1}{2}+k\right) = \frac{2k+1}{n+1},\\
y^TD^{-1}W &= \frac{2}{n+1}\left(\frac{1}{2}+\left\lfloor\frac{k}{2}\right\rfloor,\left\lceil\frac{k}{2}\right\rceil\right)\\
&= \frac{1}{n+1}\left(2\left\lfloor\frac{k}{2}\right\rfloor+1\,,\,2\left\lceil\frac{k}{2}\right\rceil\right),\\
W^TD^{-1}W &= \frac{2}{n+1}\begin{pmatrix}\frac{1}{2}+\lfloor\frac{k}{2}\rfloor & 0\\ 0 & \lceil\frac{k}{2}\rceil\end{pmatrix},\\
I_2 - W^TD^{-1}W &= \frac{1}{n+1}\begin{pmatrix}n-2\lfloor\frac{k}{2}\rfloor & 0\\ 0 & n-2\lceil\frac{k}{2}\rceil+1\end{pmatrix}.
\end{aligned}$$

Finally

$$\begin{aligned}
y^T\left[(\Upsilon)_{(1:k+1,1:k+1)}\right]^{-1}y &= \frac{2k+1}{n+1} + \frac{1}{n+1}\left(\frac{(2\lfloor k/2\rfloor+1)^2}{n-2\lfloor k/2\rfloor} + \frac{4\lceil k/2\rceil^2}{n-2\lceil k/2\rceil+1}\right)\\
&= \frac{2k+1}{n+1} + \frac{1}{n+1}\left(\frac{(k+1)^2}{n-k} + \frac{k^2}{n-k+1}\right),
\end{aligned}$$

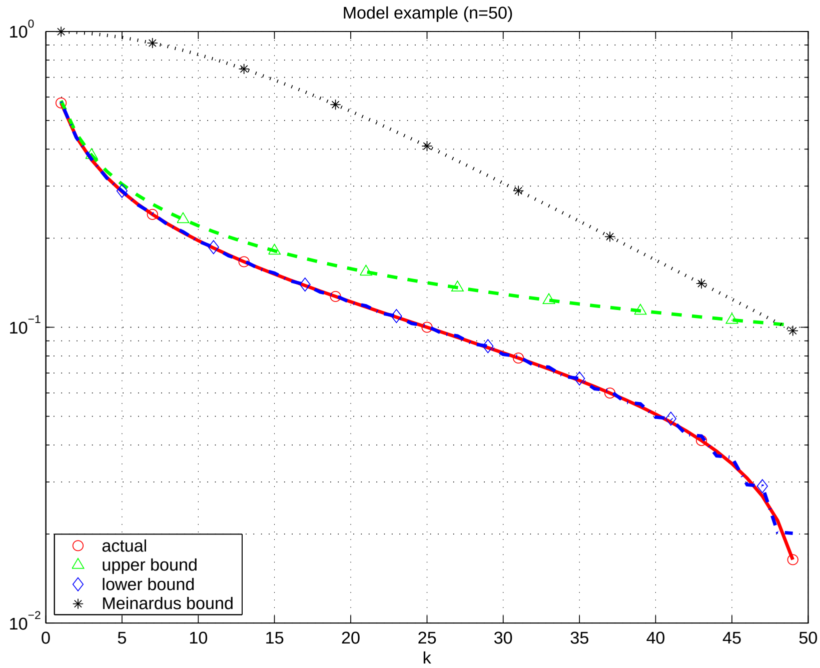

Figure 4.1: Residual $\min_{|u_{(1)}|=1} \frac{\|V_{k+1,n}^T u\|_2}{\sqrt{n}}$ *vs.* its lower and upper bounds for the model example: lower bounds nearly the same as actuals.

and hence (4.13). ∎

(4.13) has already been obtained via an direct optimization approach. (4.12) and (4.13) also show slow convergence of any method that seeks approximations from $\mathcal{K}_k(A, b)$. Figure 4.1 plots the exact value in (4.13), its lower and upper bounds by (4.12), and the upper bound (2.29) of Meinardus.

Theorem 4.2 *Let* $\Omega_n = \mathsf{diag}(\sin^2\theta_1, \ldots, \sin^2\theta_n)$*. For* $V_{k+1,n}$ *with nodes* $\alpha_j = \lambda_j$*, the eigenvalues* (4.2) *of model matrix* (4.1)*,*

$$\min_{|u_{(1)}|=1} \frac{\|\Omega_n^{1/2} V_{k+1,n}^T u\|_2}{\sqrt{(n+1)/2}} = \sqrt{\frac{6}{(2k+3)(k+2)(k+1)}}. \tag{4.14}$$

Proof: Set $p_m = U_m$, $\omega = -2$, $\tau = 1$, and $\Omega_n = \mathsf{diag}(\sin^2\theta_1, \ldots, \sin^2\theta_n)$ in (2.7). $\Upsilon \stackrel{\text{def}}{=} \boldsymbol{P}_n \Omega_n \boldsymbol{P}_n^T = \frac{n+1}{2} I$ because $\omega_j = \sin^2\theta_j$ are the scaled weights in the associated Gaussian quadrature rule [1, p.889]. Similarly to (2.20) and (2.21), we have

$$\begin{aligned} e_1^T(\bar{V}_{k+1,n}\Omega_n V_{k+1,n}^T)^{-1} e_1 &= \frac{2}{n+1} e_1^T R_{k+1} R_{k+1}^T e_1 = \frac{2}{n+1} \sum_{j=0}^{k} |U_j(1)|^2 \\ &= \frac{2}{n+1} \sum_{j=0}^{k} (j+1)^2 = \frac{2}{n+1} \frac{(2k+3)(k+2)(k+1)}{6}. \end{aligned}$$

Use the summation formula [10, p.30] to get $\sum_{j=1}^{n} \sin^2\theta_j = \frac{n+1}{2}$. (4.14) is now a consequence of (2.24). ∎

5 Concluding Remarks

This paper analyzes two examples that are hard for CG, or any method that seeks (best possible) approximate solutions within the Krylov subspace $\mathcal{K}_k(A, b)$. Frank's example has a reputation of being "the hardest case for" CG [9], and the model example is in Demmel's textbook[5] [6] to demonstrate effectiveness of various numerical methods for positive definite linear systems. Analytical formulas or extremely tight lower and upper bounds can be found for their residuals by CG or MINRES. That both problems are hard for CG is well-known, owing to their eigenvalue distribution and growing condition numbers in n. Our contribution here is quantification of the fact.

We established a general framework that could be potential useful for analyzing certain linear systems. These include the case when A's eigenvalues are $\omega(t_j - \tau)$ $(1 \le j \le n)$, where $\omega \ne 0$ and τ are two (complex) numbers, and t_j given as one of the following.

- $t_j = \cos\frac{2j-1}{2n}\pi$ $(1 \le j \le n)$, zeros of $T_n(t)$, the nth Chebyshev polynomial of the first kind. Take $p_m = T_m$ and $\Omega_n = I_n$, and then $\boldsymbol{P}_n\Omega_n\boldsymbol{P}_n^T = \frac{n}{2}\mathsf{diag}(2, 1, 1, \dots)$ [16, 17, 18].
- $t_j = \cos\frac{j-1}{n-1}\pi$ $(1 \le j \le n)$, the extreme points of $T_{n-1}(t)$ in $[-1, 1]$. Take $p_m = T_m$ and $(\Omega_n)_{(1,1)} = (\Omega_n)_{(n,n)} = 1/2$, $(\Omega_n)_{(j,j)} = 1$ for $1 < j < n$, and then $\boldsymbol{P}_n\Omega_n\boldsymbol{P}_n^T = \frac{n-1}{2}\Omega_n^{-1}$ [16, 17, 18].
- t_j $(1 \le j \le n)$ are zeros of the nth orthogonal polynomial in an orthogonal polynomial family. Take p_m to be the mth orthogonal polynomial and ω_j the weights in the associate Gaussian quadrature rule, and then $\boldsymbol{P}_n\Omega_n\boldsymbol{P}_n^T$ is diagonal [8, 17].
- $t_j = \cos\frac{2j-1}{2n+1}\pi$ $(1 \le j \le n)$ as in Section 3.
- $t_j = \exp(\iota\frac{j-1}{n}\pi)$ $(1 \le j \le n)$, where $\iota = \sqrt{-1}$. Take $p_m(t) = t^m$ and $\Omega_n = I_n$, and then $\boldsymbol{P}_n\Omega_n\boldsymbol{P}_n^T$ is diagonal.

It is possible to construct many more problems that are hard for CG or any Krylov subspace type method and justifiable by our framework. For example, if A has eigenvalues $1 + t_j$ $(1 \le j \le n)$ or $1 - t_j$ $(1 \le j \le n)$, where t_j's are zeros of $T_n(t)$, then

$$\min_{|u_{(1)}|=1} \frac{\|V_{k+1,n}^T u\|_2}{\sqrt{n}} = \sqrt{\frac{1}{2k+1}}.$$

References

[1] M. Abramowitz and I. A. Stegun (editors), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, Inc., New York, 9th printing ed., 1970.

[2] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, New York, 1996.

[3] B. Beckermann and A. B. J. Kuijlaars, *Superlinear convergence of conjugate gradients*, SIAM J. Numer. Anal., 39 (2001), pp. 300–329.

[4] ——, *Superlinear CG convergence for special right-hand sides*, Electronic Transactions on Numerical Analysis, 14 (2002), pp. 1–19.

[5] Demmel focused more on the 2-dimensional Poisson equation.

[5] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, SIAM, Philadelphia, Pennsylvania, 2nd ed., 2000.

[6] J. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[7] W. L. Frank, *Solution of linear systems by Richardson's method*, J. ACM, 7 (1960), pp. 274–286.

[8] W. Gautschi, *The condition of Vandermonde-like matrices involving orthogonal polynomials*, Linear Algebra Appl., 52/53 (1983), pp. 293–300.

[9] G. H. Golub and D. P. O'Leary, *Some history of the conjugate gradient and Lanczos algorithms: 1948-1976*, SIAM Rev., 31 (1989), pp. 50–102.

[10] I. S. Gradshteyn and I. M. Ryzhik, *Table Of Integrals, Series, and Products*, Academic Press, New York, 1980. Corrected and Enlarged Edition prepared by A. Jeffrey, incorporated the fourth edition prepared by Yu. V. Geronimus and M. Yu. Tseytlin, translaed from the RUssian by Scripta Technica, Inc.

[11] A. Greenbaum, *Comparision of splittings used with the conjugate gradient algorithm*, Numer. Math., 33 (1979), pp. 181–194.

[12] M. Hanke, *Superlinear convergence rates for the Lanczos method applied to elliptic operators*, Numer. Math., 77 (1997), pp. 487–499.

[13] I. C. F. Ipsen, *Expressions and bounds for the GMRES residual*, BIT, 40 (2000), pp. 524–535.

[14] S. Kaniel, *Estimates for some computational techniques in linear algebra*, Math. Comp., 20 (1966), pp. 369–378.

[15] R.-C. Li, *Asymptotically optimal lower bounds for the condition number of a real Vandermonde matrix*, Technical Report 2004-05, Department of Mathematics, University of Kentucky, 2004. Avaliable at `http://www.ms.uky.edu/~math/MAreport/`, and partially submitted (excluding the rectangular Vandermonde matrix part) for publications.

[16] ——, *Sharpness in rates of convergence for CG and symmetric Lanczos methods*, Technical Report 2005-01, Department of Mathematics, University of Kentucky, 2005. Avaliable at `http://www.ms.uky.edu/~math/MAreport/`.

[17] ——, *Vandermonde matrices with Chebyshev nodes*, Technical Report 2005-02, Department of Mathematics, University of Kentucky, 2005. Avaliable at `http://www.ms.uky.edu/~math/MAreport/`, and submitted.

[18] ——, *On Meinardus' examples for the conjugate gradient method*, Technical Report 2005-06, Department of Mathematics, University of Kentucky. Avaliable at `http://www.ms.uky.edu/~math/MAreport/`, and submitted.

[19] J. Liesen, M. Rozložník, and Z. Strakos, *Least squares residuals and minimal residual methods*, SIAM J. Sci. Comput., 23 (2002), pp. 1503–1525.

[20] J. Liesen and P. Tichý, *The worst-case GMRES for normal matrices*, BIT, 44 (2004), pp. 79–98.

[21] G. Meinardus, *Über eine Verallgemeinerung einer Ungleichung von L. V. Kantorowitsch*, Numer. Math., 5 (1963), pp. 14–23.

[22] A. E. Naiman, I. M. Babuška, and H. C. Elman, *A note on conjugate gradient convergence*, Numer. Math., 76 (1997), pp. 209–230.

[23] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.

[24] Y. Saad, *On the rates of convergence of the Lanczos and the block-Lanczos methods*, SIAM J. Numer. Anal., 15 (1980), pp. 687–706.

[25] G. L. G. Sleijpen and A. van der Sluis, *Further results on the convergence behavior of conjugate-gradients and Ritz values*, Linear Algebra Appl., 246 (1996), pp. 233–378.

[26] A. van der Sluis and H. A. van der Vorst, *The convergence behavior of Ritz values in the presence of close eigenvalues*, Linear Algebra Appl., 88/89 (1987), pp. 651–694.

[27] D. Young, *Computation of eigenvalues and eigenvectors of a certain matrix*, Numerical Analysis Notes NN-18, Space Technology Laboratories, August 1956.