ON THE ROOTS OF THE ORTHOGONAL POLYNOMIALS AND RESIDUAL POLYNOMIALS ASSOCIATED WITH A CONJUGATE GRADIENT METHOD

THOMAS MANTEUFFEL and JAMES OTTO

Department of Mathematics
University of Colorado at Denver
Denver, CO 80204.

ABSTRACT

In this paper we explore two sets of polynomials, the orthogonal polynomials and the residual polynomials, associated with a preconditioned conjugate gradient iteration for the solution of the linear system $A\underline{x} = \underline{b}$. In the context of preconditioning by the matrix C, we show that the roots of the orthogonal polynomials, also known as generalized Ritz values, are the eigenvalues of an orthogonal section of the matrix CA while the roots of the residual polynomials, also known as pseudo-Ritz values (or roots of kernel polynomials), are the reciprocals of the eigenvalues of an orthogonal section of the matrix $(CA)^{-1}$. When CA is self-adjoint positive definite, this distinction is minimal, but for the indefinite or non-self-adjoint case this distinction becomes important. We use these two sets of roots to form possibly non-convex regions in the complex plane that describe the spectrum of CA.

Key words: orthogonal polynomials, residual polynomials, conjugate gradient method, Ritz values, field of values

1. Introduction

1.1. Motivation

A preconditioned conjugate gradient method for the solution of the $(n \times n)$ linear system

$$A\underline{x} = \underline{b}, \tag{1.1}$$

with preconditioning matrix C is defined by the additional choice of an inner product which we denote by $\langle B\cdot, \cdot\rangle$, where B is an Hermitian positive definite (HPD) matrix. Having chosen B, C, A and an initial guess $\underline{x}_0$, and thus, initial residual $\underline{r}_0 = \underline{b} - A\underline{x}_0$, the subsequent iterates are uniquely determined by the two hypotheses:

$$\begin{aligned} \underline{x}_{i+1} &= \underline{x}_i + \underline{z} \qquad \underline{z} \in \mathcal{K}_{i+1}(C\underline{r}_0, CA), && (1.2)\\ \underline{e}_{i+1} &\perp_B \underline{z} \qquad \forall \underline{z} \in \mathcal{K}_{i+1}(C\underline{r}_0, CA). && (1.3)\end{aligned}$$

Here,

$$\mathcal{K}_{i+1}(C\underline{r}_0, CA) = sp\{C\underline{r}_0, CAC\underline{r}_0, \ldots, (CA)^i C\underline{r}_0\} \tag{1.4}$$

is the Krylov subspace of dimension $i+1$ generated by $C\underline{r}_0$ and CA, $\underline{e}_{i+1} = \underline{x} - \underline{x}_{i+1}$ is the error at step $i+1$, and $\perp_B$ represents orthogonality in the B-inner product (c.f. [AMS90]). Several algorithms exist for implementing this method (see [YoJe80], [JeYo83], [SaSc86], [AMS90]). To start we will consider the Orthodir algorithm. Orthomin will be discussed in Section 3.2. Given $\underline{x}_0$, the iteration is

$$\begin{aligned}
&\underline{r}_0 = \underline{b} - A\underline{x}_0, && && (1.5a)\\
&\underline{s}_0 = C\underline{r}_0, && && (1.5b)\\
&\underline{p}_0 = \underline{s}_0, && && (1.5c)\\
&\vdots \\
&\underline{x}_{j+1} = \underline{x}_j + \alpha_j \underline{p}_j, && \alpha_j = \frac{\langle B\underline{e}_j, \underline{p}_j\rangle}{\langle B\underline{p}_j, \underline{p}_j\rangle}, && (1.5d)\\
&\underline{r}_{j+1} = \underline{r}_j - \alpha_j A\underline{p}_j, && && (1.5e)\\
&\underline{s}_{i+1} = C\underline{r}_{i+1}, && && (1.5f)\\
&\underline{p}_{j+1} = CA\underline{p}_j - \textstyle\sum_{i=0}^{j} \sigma_{ij}\underline{p}_i, && \sigma_{ij} = \frac{\langle BCA\underline{p}_j, \underline{p}_i\rangle}{\langle B\underline{p}_i, \underline{p}_i\rangle}. && (1.5g)
\end{aligned}$$

Here we have introduced the preconditioned residual $\underline{s}_i = C\underline{r}_i$.

This algorithm yields a three-term recursion for $\underline{p}_i$ for every choice of $\underline{x}_0$ if and only if CA is B-normal(1) [FaMa84]. In this paper we are interested in the general situation and will consider the existence of a three-term recursion as a special case.

The properties (1.2) and (1.3) above lead to the relationships

$$\begin{aligned} \langle B\underline{p}_i, \underline{p}_j\rangle &= 0 \quad i \neq j, && (1.6)\\ \langle B\underline{e}_i, \underline{s}_j\rangle &= 0 \quad i > j. && (1.7)\end{aligned}$$

For a more detailed presentation of these properties see [AMS90].

There are two sets of polynomials associated with this iteration. From (1.5) it is easy to see that

$$\underline{s}_i, \underline{p}_i \in \mathcal{K}_{i+1}(\underline{s}_0, CA). \tag{1.8}$$

Thus, we can express $\underline{s}_i$ and $\underline{p}_i$ as polynomials of degree i in CA times $\underline{s}_0$, that is,

$$\begin{aligned} \underline{s}_i &= \rho_i(CA)\underline{s}_0, && (1.9) \\ \underline{p}_i &= \phi_i(CA)\underline{s}_0. && (1.10) \end{aligned}$$

The polynomials $\rho_i(\lambda)$ and $\phi_i(\lambda)$ are called the residual polynomials and the B-orthogonal polynomials respectively. For brevity, the B will be suppressed when the meaning is obvious. In this paper we will examine the properties of the roots of these two sets of polynomials. Both have been used to approximate eigenvalues of the matrix CA. The roots of the orthogonal polynomials have been referred to as Ritz values when $B = I$ (c.f. [Pa80]). We will refer to them as generalized Ritz values or B-Ritz values or just Ritz values when the choice of B is clear. In Section 2 we will show that the roots of the orthogonal polynomials are the eigenvalues of a B-orthogonal section of CA onto the subspace $\mathcal{K}_i(\underline{s}_0, CA)$ and, thus, lie in the B-field of values of CA (see Section 2.1). The residual polynomials have been referred to as kernel polynomials [St58] and their roots have been referred to as pseudo-Ritz values [Fr89]. The roots of the residual polynomials will be shown to be the reciprocals of the eigenvalues of a B-orthogonal section of $(CA)^{-1}$ onto $\mathcal{K}_i(\underline{s}_0, CA)$. Thus, the reciprocals of the roots of the residual polynomials will lie in the B-field of values of $(CA)^{-1}$ (see Section 2.2). If CA is B-self-adjoint positive definite there is no difference between the B-field of values of CA and the reciprocals of the B-field of values of $(CA)^{-1}$. However, in the indefinite and non-self-adjoint case the distinction is important. In Section 3 we will show that both sets of roots can be generated from essentially the same data which can be accumulated during the iteration at a small additional cost.

The motivation for this paper is the desire to describe the spectrum of CA when CA is not B-self-adjoint positive definite by a possibly non-convex set in the complex plane $\mathcal{C}$. Such a set would be useful in constructing a polynomial preconditioning [JMP83], [Sa83], [Sa85], [AMS89], [As91], [AMO91] or in constructing a Richardson-type iteration [SmSa82], [NRT92]. This will be the topic of Section 4. Section 5 will contain numerical results.

Although this paper is focused on the conjugate gradient method, many of the results apply in a more general setting. In particular, they apply to the QMR method of [FrNa91]. A conjugate gradient method fixes the inner product matrix B for all $\underline{r}_0$. This is the basis for the results in [FaMa84] and [FaMa87]. The QMR method also satisfies (1.2) and (1.3), but for a B that depends on $\underline{r}_0$. The QMR method generates a series of step directions, $\underline{p}_i$, and preconditioned residuals, $\underline{s}_i$, that satisfy (1.6) and (1.7) for this particular B. Since the results of Section 2.1 and Section 2.2 depend only upon (1.6) and (1.7), they apply to the QMR method. The residual polynomials associated with the QMR method are studied in [Fr91] where they are referred to as quasi-kernel polynomials.

1.2. Notation

Throughout this paper we will employ the following notational conventions. Particular usage should become clear from the context.

$\mathcal{R}^n, \mathcal{C}^n$	–	Vector spaces of real and complex $n-$tuples.
$\mathcal{R}^{n\times m}, \mathcal{C}^{n\times m}$	–	Vector spaces of real and complex $n \times m$ matrices.
$\mathcal{P}_k$	–	Space of polynomials of degree at most k.
$\mathcal{K},\ \mathcal{V}, \ldots$	–	Other calligraphic letters denote subspaces of $\mathcal{R}^n$ or $\mathcal{C}^n$.
$A, B, \ldots$	–	Upper case Roman and Greek letters denote matrices.
$\underline{x}, \underline{\xi}, \ldots$	–	Underlined lower case Roman and Greek letters denote vectors.
$a, \alpha, \ldots$	–	Lower case Roman and Greek letters denote scalars.
$\langle \cdot, \cdot \rangle$	–	Euclidean inner product on $\mathcal{C}^n$.
$\langle B\cdot, \cdot \rangle$	–	$B-$inner product on $\mathcal{C}^n$.
A^*	–	Euclidean adjoint of A.
$A^\dagger$	–	$B-$adjoint of A.
$sp\{\underline{x}_j\}$	–	The linear span of the vectors $\underline{x}_j$.
$\Sigma(A)$	–	Spectrum of A.
$\mathcal{H}(A)$	–	Convex hull of $\Sigma(A)$.
$\mathcal{F}_B(A)$	–	$B-$Field of values of A.

2. The Roots

In this section we will examine some properties of the roots of the two sets of polynomials, beginning with the orthogonal polynomials.

2.1. Orthogonal Polynomials

The orthogonal polynomials (1.10) are characterized by the condition (1.6), that is, $\underline{p}_i$ is the unique (up to scale) vector such that

$$\underline{p}_i \in \mathcal{K}_{i+1}(\underline{s}_0, CA), \tag{2.1}$$

$$\underline{p}_i \perp_B \mathcal{K}_i(\underline{s}_0, CA). \tag{2.2}$$

Notice that (1.5g) implies that $\underline{p}_i = \phi_i(CA)\underline{s}_0$ where $\phi_i(\lambda)$ is of exact degree i. The following is a generalization of a well known result for $B = C = I$ (c.f. [Pa80]).

Theorem 2.1. Suppose that $\underline{p}_i = \phi_i(CA)\underline{s}_0 \neq \underline{0}$ with $\phi_i(\lambda) = \prod_{j=1}^{b_i}(\lambda - \lambda_j)^{m_j}$ and $d_i \equiv \sum_{j=1}^{b_i} m_j = i$. Suppose

$$\underline{p}_i \perp_B \mathcal{K}_i(\underline{s}_0, CA). \tag{2.3}$$

Let V_i and W_i be any two $n \times i$ matrices whose columns span $\mathcal{K}_i(\underline{s}_0, CA)$. Then the roots of

$$\phi_i(\lambda) = 0 \tag{2.4}$$

are the solutions of the generalized eigenvalue problem

$$W_i^* B(CA)V_i\underline{w} = \lambda W_i^* BV_i\underline{w}. \tag{2.5}$$

If $m_j > 1$, then $(W_i^* BV_i)^{-1}(W_i^* B(CA)V_i)$ has a Jordan block of size m_j associated with the eigenvalue λ_j.

Proof. Suppose λ_j is a root of $\phi_i(\lambda) = 0$. We may write

$$\underline{p}_i = (CA - \lambda_j I)\underline{q}_1 = (CA - \lambda_j I)^2\underline{q}_2 = \ldots = (CA - \lambda_j I)^{m_j}\underline{q}_{m_j} \tag{2.6}$$

with $\underline{q}_k \in \mathcal{K}_{i-k+1}(\underline{s}_0, CA)$ for $k = 1, \ldots, m_j$. The vectors $\underline{q}_1, \ldots, \underline{q}_{m_j}$ are linearly independent. If not, then $\underline{p}_i \in \mathcal{K}_i(\underline{s}_0, CA)$, which contradicts the hypothesis. Let $\underline{q}_k = V_i\underline{w}_k$ for $k = 1, \ldots, m_j$. The vectors $\underline{w}_1, \ldots, \underline{w}_{m_j}$ are independent by the same argument. By (2.3) we have

$$\langle B(CA - \lambda_j I)V_i\underline{w}_1, W_i\underline{z}\rangle_{\mathcal{C}^n} = 0, \quad \forall \underline{z} \in \mathcal{C}^i, \tag{2.7}$$

or

$$\langle (W_i^* BCAV_i - \lambda_j W_i^* BV_i)\underline{w}_1, \underline{z}\rangle_{\mathcal{C}^i} = 0, \quad \forall \underline{z} \in \mathcal{C}^i, \tag{2.8}$$

or

$$W_i^* BCAV_i\underline{w}_1 = \lambda_j W_i^* BV_i\underline{w}_1, \tag{2.9}$$

which establishes (2.5). Now, we also have

$$V_i\underline{w}_k = (CA - \lambda_j I)V_i\underline{w}_{k+1} \tag{2.10}$$

or

$$W_i^* BV_i\underline{w}_k = W_i^* BCAV_i\underline{w}_{k+1} - \lambda_j W_i^* BV_i\underline{w}_{k+1} \tag{2.11}$$

for $k = 1, \ldots, m_j - 1$. Since $W_i^* BV_i$ is nonsingular, $\underline{w}_1, \ldots, \underline{w}_{m_j}$ form a basis for a nondiagonal Jordan block associated with λ_j. $\square$

Let us rewrite (2.5) as

$$V_i(W_i^* BV_i)^{-1}W_i^* B(CA)V_i\underline{w} = \lambda V_i\underline{w}, \tag{2.12}$$

and recognize the operator on the left as the restriction of CA to $\mathcal{K}_i$ followed by the B-orthogonal projection back onto $\mathcal{K}_i$. Thus, the solutions of (2.5) are the eigenvalues of the B-orthogonal section of CA onto $\mathcal{K}_i$ and $\phi_i(\lambda)$ is the minimal polynomial of the B-orthogonal section. We now show that the roots of $\phi_i(\lambda) = 0$ are in the B-field of values of CA. First, we recall some definitions. The B-adjoint of a matrix is the adjoint with respect to the B-inner product, that is, $A^\dagger$ is the unique matrix such that

$$\langle BA\underline{x}, \underline{y}\rangle = \langle B\underline{x}, A^\dagger \underline{y}\rangle \quad \forall \underline{x},\ \underline{y} \neq \underline{0} \in \mathcal{C}^n. \tag{2.13}$$

We say a matrix is B-normal if it is normal with respect to the B-inner product, that is, if one of the following equivalent properties holds (c.f. [Ga77], [FaMa84]):

$$\begin{array}{rl} i) & A^\dagger A = AA^\dagger. \\ ii) & A \text{ and } A^\dagger \text{ have the same complete set of } \perp_B-\text{eigenvectors}. \\ iii) & A^\dagger = q_s(A) \text{ where } q_s \text{ is a polynomial of degree } s. \end{array} \tag{2.14}$$

If s is the smallest degree for which the third property holds we say that A is of normal degree s or is B-normal(s). Finally, recall that the B-field of values of A is defined as

$$\mathcal{F}_B(A) = \{\lambda : \lambda = \frac{\langle BA\underline{x}, \underline{x}\rangle}{\langle B\underline{x}, \underline{x}\rangle} \text{ for some } \underline{x} \neq \underline{0} \in \mathcal{C}^n\}. \tag{2.15}$$

The B-field of values has the following properties (c.f. [Ho64], [HoJo91], [Ei91]).

$$\begin{array}{rl} i) & \mathcal{F}_B(A) \text{ is a convex set in } \mathcal{C}. \\ ii) & \Sigma(A) \subseteq \mathcal{H}(A) \subseteq \mathcal{F}_B(A). \\ iii) & \mathcal{H}(A) = \mathcal{F}_B(A) \text{ if } A \text{ is } B-\text{normal}. \end{array} \tag{2.16}$$

The difference between $\mathcal{H}(A)$ and $\mathcal{F}_B(A)$ has been used as a measure of the B-normality of A (c.f. [He62], [Ta68]).

Corrollary 2.2. If λ is a root of an orthogonal polynomial then

$$\lambda \in \mathcal{F}_B(CA). \tag{2.17}$$

Proof. Let $W_i = V_i$ and $\underline{z} = \underline{w}_1 = \underline{w}$ in (2.7). We have

$$\lambda = \frac{\langle BCAV_i\underline{w}, V_i\underline{w}\rangle}{\langle BV_i\underline{w}, V_i\underline{w}\rangle}, \tag{2.18}$$

which shows that if $\phi_i(\lambda) = 0$, then $\lambda \in \mathcal{F}_B(CA)$. $\square$

With the proper choice of of $\underline{x}_0$, any value within $\mathcal{F}_B(CA)$ can be obtained as a root of the orthogonal polynomial of degree 1. From (1.3) and (1.5g) we see

$$\phi_1(\lambda) = \lambda - \frac{\langle BCA\underline{p}_0, \underline{p}_0\rangle}{\langle B\underline{p}_0, \underline{p}_0\rangle}. \tag{2.19}$$

Since $\underline{p}_0 = \underline{s}_0$ can be chosen arbitrarily, the root of the first orthogonal polynomial can take any value in $\mathcal{F}_B(CA)$.

2.2. Residual Polynomials

We first remark that the residual polynomials (1.9) are normalized at the origin, that is, $\rho_i(0) = 1$. This can be seen from the recursion (1.5e) and (1.5f) which yield $\rho_i(0) = \rho_{i-1}(0) = \cdots = \rho_0(0) = 1$. The residual polynomials are characterized by (1.7). To see this, notice that we may write $\underline{e}_i = (CA)^{-1}\underline{s}_i$. Thus, $\underline{s}_i$ can be characterized by

$$\begin{aligned} \underline{s}_i &\in \mathcal{K}_{i+1}(\underline{s}_0, CA), && (2.20)\\ (CA)^{-1}\underline{s}_i &\perp_B \mathcal{K}_i(\underline{s}_0, CA). && (2.21)\end{aligned}$$

An alternative characterization will shed some light on the main result of this section. Suppose we define $\underline{u}_i = (CA)^i\underline{s}_0$. Then we have

$$\begin{aligned}\mathcal{K}_i(\underline{s}_0, CA) &= sp\{\underline{s}_0, CA\underline{s}_0, \ldots, (CA)^{(i-1)}\underline{s}_0\}\\ &= sp\{(CA)^{-(i-1)}\underline{u}_{i-1}, \ldots, (CA)^{-1}\underline{u}_{i-1}, \underline{u}_{i-1}\}\\ &= \mathcal{K}_i(\underline{u}_{i-1}, (CA)^{-1}).\end{aligned}$$

In addition, we have

$$\begin{aligned}(CA)^{-1}\mathcal{K}_{i+1}(\underline{s}_0, CA) &= sp\{\underline{e}_0, \underline{s}_0, CA\underline{s}_0, \ldots, (CA)^{(i-1)}\underline{s}_0\}\\ &= sp\{(CA)^{-i}\underline{u}_{i-1}, \ldots, (CA)^{-1}\underline{u}_{i-1}, \underline{u}_{i-1}\}\\ &= \mathcal{K}_{i+1}(\underline{u}_{i-1}, (CA)^{-1}).\end{aligned}$$

Finally, we use the expression

$$\underline{e}_i = (CA)^{-1}\underline{s}_i = (CA)^{-1}\rho_i(CA)\underline{s}_0 = \rho_i(CA)(CA)^{-i}\underline{u}_{i-1}, \tag{2.22}$$

and reformulate (2.20) and (2.21) as

$$\begin{aligned}\underline{e}_i &\in \mathcal{K}_{i+1}(\underline{u}_{i-1}, (CA)^{-1}), && (2.23)\\ \underline{e}_i &\perp_B \mathcal{K}_i(\underline{u}_{i-1}, (CA)^{-1}). && (2.24)\end{aligned}$$

This motivates the definition

$$\hat{\phi}_i(\mu) \equiv \mu^i \rho_i(\frac{1}{\mu}). \tag{2.25}$$

From (2.22) we have

$$\underline{e}_i = \hat{\phi}_i((CA)^{-1})\underline{u}_{i-1}. \tag{2.26}$$

Together with (2.23) and (2.24) we see that $\hat{\phi}_i(\mu)$ is the orthogonal polynomial of degree i generated by $\underline{u}_{i-1}$ and $(CA)^{-1}$ and, thus, is the minimal polynomial of the B-orthogonal section of $(CA)^{-1}$ onto $\mathcal{K}_i$. We now prove the main result of this section.

Theorem 2.3. Suppose that $\underline{s}_i = \rho_i(CA)\underline{s}_0 \neq 0$ with $\rho_i(\lambda) = \prod_{j=1}^{b_i}(1 - \frac{\lambda}{\lambda_j})^{m_j}$ and $d_i \equiv \sum_{j=1}^{b_i} m_j \leq i$. Suppose

$$(CA)^{-1}\underline{s}_i \perp_B \mathcal{K}_i(\underline{s}_0, CA). \tag{2.27}$$

Let V_i and W_i be any two $(n \times i)$ matrices whose columns span $\mathcal{K}_i(\underline{s}_0, CA)$. If λ is a root of

$$\rho_i(\lambda) = 0 \tag{2.28}$$

then $\frac{1}{\lambda}$ is a solution of the generalized eigenvalue problem

$$W_i^* B(CA)^{-1} V_i \underline{w} = \frac{1}{\lambda} W_i^* B V_i \underline{w}. \tag{2.29}$$

If $m_j > 1$, then $(W_i^* BV_i)^{-1}(W_i^* B(CA)^{-1}V_i)$ has a Jordan block of size m_j associated with the eigenvalue $\frac{1}{\lambda_j}$. If $d_i < i$, then $(W_i^* BV_i)^{-1}(W_i^* B(CA)^{-1}V_i)$ has a nondiagonal Jordan block of size $i - d_i$ associated with a zero eigenvalue.

Proof. Suppose λ_j is a root of $\rho_i(\lambda) = 0$. We may write

$$\underline{s}_i = (I - \frac{1}{\lambda_j} CA)\underline{q} \tag{2.30}$$

with $\underline{q} \in \mathcal{K}_i(\underline{s}_0, CA)$. Let $\underline{q} = V_i \underline{w}$. Then,

$$\langle B(CA)^{-1}(I - \frac{1}{\lambda_j} CA)V_i \underline{w}, W_i \underline{z} \rangle_{\mathcal{C}^n} = 0, \quad \forall \underline{z} \in \mathcal{C}^i, \tag{2.31}$$

or

$$\langle (W_i^* B(CA)^{-1} V_i - \frac{1}{\lambda_j} W_i^* B V_i)\underline{w}, \underline{z} \rangle_{\mathcal{C}^i} = 0, \quad \forall \underline{z} \in \mathcal{C}^i, \tag{2.32}$$

or

$$W_i^* B(CA)^{-1} V_i \underline{w} = \frac{1}{\lambda_j} W_i^* B V_i \underline{w}, \tag{2.33}$$

which establishes (2.29). The remaining results can be obtained by the observation that $\hat{\phi}_i(\mu) = \mu^i \rho_i(\frac{1}{\mu})$ is the orthogonal polynomial of degree i generated from $\underline{u}_{i-1}$ and $(CA)^{-1}$. We have

$$\hat{\phi}_i(\mu) = \mu^{(i-d_i)} \prod_{j=1}^{b_i} (\mu - \frac{1}{\lambda_j})^{m_j}. \tag{2.34}$$

An application of Theorem 2.1 yields the result. $\square$

Corrollary 2.4. If λ is a root of a residual polynomial then

$$\frac{1}{\lambda} \in \mathcal{F}_B((CA)^{-1}). \tag{2.35}$$

Proof. As in the proof of Corollary 2.1, we let $W_i = V_i$ and $\underline{z} = \underline{w}$ in (2.31) to get

$$\frac{1}{\lambda} = \frac{\langle B(CA)^{-1} V_i \underline{w}, V_i \underline{w} \rangle}{\langle B V_i \underline{w}, V_i \underline{w} \rangle}, \tag{2.36}$$

which shows that if $\rho_i(\lambda) = 0$, then $\frac{1}{\lambda} \in \mathcal{F}_B((CA)^{-1})$. $\square$

Let us define the set

$$\mathcal{F}_B^{-1}((CA)^{-1}) = \{\lambda : \frac{1}{\lambda} \in \mathcal{F}_B((CA)^{-1})\}. \tag{2.37}$$

With the proper choice of $\underline{x}_0$, any value within the set (2.37) can be obtained as a root of the residual polynomial of degree 1. From (1.5e) and (1.5f) we see that the first residual polynomial is

$$\rho_1(\lambda) = 1 - \frac{\langle B\underline{e}_0, \underline{p}_0 \rangle}{\langle B\underline{p}_0, \underline{p}_0 \rangle} \lambda. \tag{2.38}$$

From (1.5c), $\underline{p}_0 = \underline{s}_0$, and $\underline{e}_o = (CA)^{-1}\underline{s}_0$, we see that $\rho_1(\lambda) = 0$ implies

$$\frac{1}{\lambda} = \frac{\langle B(CA)^{-1}\underline{s}_0, \underline{s}_0 \rangle}{\langle B\underline{s}_0, \underline{s}_0 \rangle} \tag{2.39}$$

and thus, λ can take on any value in $\mathcal{F}_B^{-1}((CA)^{-1})$.

2.3. Implications

The Ritz values of the matrix A with respect to the subspace W are generally understood to be the eigenvalues of the orthogonal section of A onto W in the standard inner product (c.f. [Pa80]). Theorem 2.1 shows that the roots of the B-orthogonal polynomials are the eigenvalues of a B-orthogonal section of CA onto $\mathcal{K}_i(\underline{s}_0, CA)$. In this paper these roots will be referred to as generalized Ritz values or B-Ritz values or just Ritz values when the B in question is clear. In [Fr89] the roots of the residual polynomials are referred to as the pseudo-Ritz values of CA with respect to $\mathcal{K}_i(\underline{s}_0, CA)$. Theorem 2.3 shows that they are the reciprocals of the nonzero B-Ritz values of $(CA)^{-1}$ with respect to $\mathcal{K}_i(\underline{s}_0, CA)$.

Clearly, the choice of B makes a difference. When the Lanczos algorithm is applied to a Hermitian matrix, the resulting eigenvalue estimates are Ritz values ($B = I$). The relationship between the conjugate gradient method and the Lanczos algorithm has been described in a variety of papers and used to obtain eigenvalue estimates (cf. [CGO76]). However, in the general case it is the roots of the orthogonal polynomials not the residual polynomials that yield B-Ritz values (c.f. [AMS90]). For many implementations the distinction is minimal. For example, consider the original algorithm of Hestenes and Stiefel (CGHS) [HeSt52]. Here A is assumed to be HPD. $C = I$ and $B = A$. It is easy to see that (1.6) and (1.7) become

$$\underline{r}_i \in \mathcal{K}_{i+1}(\underline{r}_0, A), \tag{2.40}$$

$$\underline{r}_i \perp \mathcal{K}_i(\underline{r}_0, A). \tag{2.41}$$

and the roots of the residual polynomials are in fact the Ritz values ($B = I$) of A with respect to $\mathcal{K}_i(\underline{r}_0, A)$, while the roots of the orthogonal polynomials are B-Ritz values ($B = A$) of A with respect to $\mathcal{K}_i(\underline{r}_0, A)$.

In the preconditioned conjugate gradient algorithm (PCG) [CGO76] the matrices A and C are assumed to be HPD with $B = A$. Under these hypotheses, CA is B-self-adjoint positive definite, that is,

$$\langle BCA\underline{x}, \underline{y}\rangle = \langle B\underline{x}, CA\underline{y}\rangle, \tag{2.42}$$

for every $\underline{x}, \underline{y} \neq \underline{0} \in \mathcal{C}^n$ and

$$\frac{\langle BCA\underline{x}, \underline{x}\rangle}{\langle B\underline{x}, \underline{x}\rangle} > 0 \tag{2.43}$$

for every $\underline{x} \in \mathcal{C}^n$. While the roots of the residual polynomials generated by PCG are not B-Ritz values of CA with respect to $\mathcal{K}_i(\underline{s}_0, CA)$, they are the Ritz values ($B = I$) of $C^{1/2}AC^{1/2}$ with respect to $\mathcal{K}_i(C^{1/2}\underline{r}_0, C^{1/2}AC^{1/2})$. We have a more general result:

Lemma 2.5. If CA is B-self-adjoint positive definite then the roots of the residual polynomials are the eigenvalues of the $B(CA)^{-1}$-orthogonal section of CA onto $\mathcal{K}_i(\underline{s}_0, CA)$.

Proof. We first show that BCA and $B(CA)^{-1}$ are HPD. From (2.42) we have

$$\begin{aligned} BCA &= (CA)^*B = (BCA)^*, & (2.44) \\ B(CA)^{-1} &= (CA)^{-*}B = (B(CA)^{-1})^*. & (2.45) \end{aligned}$$

which shows that BCA and $B(CA)^{-1}$ are Hermitian. Consider

$$\frac{\langle BCA\underline{x}, \underline{x}\rangle}{\langle \underline{x}, \underline{x}\rangle} = \frac{\langle BCA\underline{x}, \underline{x}\rangle}{\langle B\underline{x}, \underline{x}\rangle}\frac{\langle B\underline{x}, \underline{x}\rangle}{\langle \underline{x}, \underline{x}\rangle}. \tag{2.46}$$

Since B is HPD, (2.43) implies that BCA is HPD. The proof for $B(CA)^{-1}$ is similar.

The final result follows from (2.20) and (2.21) which can be rewritten as

$$\begin{aligned} \underline{s}_i &\in \mathcal{K}_{i+1}(\underline{s}_0, CA), & (2.47) \\ \underline{s}_i &\perp_{B(CA)^{-1}} \mathcal{K}_i(\underline{s}_0, CA). & (2.48) \end{aligned}$$

This completes the proof. $\square$

Thus, for PCG the roots of the residual polynomials are the eigenvalues of a C^{-1}-orthogonal section. In terms of fields of values we have the following result.

Lemma 2.6. If CA is B-normal, then the roots of the orthogonal polynomials lie in $\mathcal{H}(CA)$. If CA is B-self-adjoint positive definite then the roots of the residual polynomials lie in $\mathcal{H}(CA)$.

Proof. When CA is B-normal, we see from (2.16) that $\mathcal{F}_B(CA) = \mathcal{H}(CA)$. Corollary 2.2 implies that the roots of the orthogonal polynomials are in $\mathcal{H}(CA)$.

Now, by the same reasoning, $\mathcal{F}_B((CA)^{-1}) = \mathcal{H}((CA)^{-1})$. If CA is B-self-adjoint positive definite, then both $\mathcal{H}(CA)$ and $\mathcal{H}((CA)^{-1})$ are intervals on the positive real line. In fact, if $\frac{1}{\lambda} \in \mathcal{H}((CA)^{-1})$ then $\lambda \in \mathcal{H}(CA)$. Corollary 2.4 completes the proof. $\square$

If CA is not B-self-adjoint positive definite, then the roots of the residual polynomials may not lie in $\mathcal{H}(CA)$. We offer two examples. First, suppose CA is B-self-adjoint but indefinite. Then $(CA)^{-1}$ is also indefinite. Suppose $\Sigma(CA) \subseteq [a,b] \cup [c,d]$ with $b < 0 < c$. Then $\mathcal{F}_B(CA) = \mathcal{H}(CA) = [a,b]$, while $\mathcal{F}_B((CA)^{-1}) = \mathcal{H}((CA)^{-1}) = [\frac{1}{b}, \frac{1}{c}]$ contains the origin. The set $\mathcal{F}_B^{-1}((CA)^{-1}) = (-\infty, b] \cup [c, \infty)$ contains the entire real line except the interval (b,c). We will show in Corollary 3.6 that, in this case, each orthogonal polynomial may have at most one root inside (b,c), while each residual polynomial may have at most one root outside $[a,d]$.

Next, consider the case when CA is B-normal(1) (the most general case in which the conjugate gradient algorithm can be implemented with a three term recursion) but not B-self-adjoint. It was shown in [FaMa84] that if CA is B-normal(1) we may write

$$CA = e^{i\theta}(G + irI) \tag{2.49}$$

where $r \in \mathcal{R}$ and G is B-self-adjoint. The eigenvalues of G lie in an interval on the real line, say $[a, b]$. This implies that $\Sigma(CA)$ lies in the line segment $[e^{i\theta}(a+ir), e^{i\theta}(b+ir)]$ in $\mathcal{C}$ (see Fig 1.). From (2.16) we again see that $\mathcal{F}_B(CA) = \mathcal{H}(CA)$ and thus, the roots of the orthogonal polynomials will lie on the line segment.

Consider $\Sigma((CA)^{-1})$. If $\lambda \in \Sigma(CA)$ then $\lambda = e^{i\theta}(\mu + ir)$ for some $\mu \in [a, b]$. Thus, $\Sigma((CA)^{-1})$ is contained in the curve defined by

$$\lambda = e^{i\theta}\frac{\mu - ir}{\mu^2 + r^2} \quad \text{for} \quad \mu \in [a, b]. \tag{2.50}$$

(see Figure 1). This linear fractional transformation maps an interval onto either an arc of a circle or an interval (possibly infinite). Since $\Sigma((CA)^{-1})$ is no longer contained in a line segment $\mathcal{F}_B((CA)^{-1}) = \mathcal{H}((CA)^{-1})$ contains points not on the curve (2.50). Likewise, the set $\mathcal{F}_B^{-1}((CA)^{-1})$ is larger than $\mathcal{F}_B(CA)$ (see Figure 1). In the case when CA is not B-normal the discrepancy may be even greater (see Section 5).

3. Computation of the Roots

In this section we will show that the roots of the orthogonal polynomials and the roots of the residual polynomials can be computed from essentially the same information accumulated during a conjugate gradient iteration at a small additional cost. The development is similar to the matrix analysis in [AsGu91].

3.1. The General Case

Let us begin by considering the matrix form of the recursions in (1.5). Suppose we let P_j be the $n \times j$ matrix whose columns are the $\underline{p}_i$'s. That is, let

$$P_j = \begin{bmatrix} | & | & & | \\ \underline{p}_0 & \underline{p}_1 & \cdots & \underline{p}_{j-1} \\ | & | & & | \end{bmatrix}, \tag{3.1}$$

and define S_j similarly. From (1.6) we have

$$P_j^* B P_j = \Delta_j = diag(\delta_0, \ldots, \delta_{j-1}), \tag{3.2}$$

where $\delta_i = \langle B\underline{p}_i, \underline{p}_i \rangle$. The recurrence in (1.5g) can be expressed in matrix form as

$$CAP_j = P_{j+1} \widetilde{H}_{j+1} \tag{3.3}$$

where

$$\widetilde{H}_{j+1} = \begin{bmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0,j-1} \\ 1 & \sigma_{11} & \cdots & \sigma_{1,j-1} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{j-1,j-1} \\ 0 & \ldots & 0 & 1 \end{bmatrix}_{j+1 \times j} . \tag{3.4}$$

Together with (3.2), we have

$$P_j^* BCAP_j = \Delta_j H_j, \tag{3.5}$$

where H_j is derived from $\widetilde{H}_{j+1}$ by deleting the last row; that is,

$$H_j = \begin{bmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0,j-1} \\ 1 & \sigma_{11} & \cdots & \sigma_{1,j-1} \\ & \ddots & \ddots & \vdots \\ & & 1 & \sigma_{j-1,j-1} \end{bmatrix}_{j \times j} . \tag{3.6}$$

Theorem 3.1. The eigenvalues of H_j are the roots of the orthogonal polynomial, $\phi_j(\lambda) = 0$.

Proof. With $V_j = W_j = P_j$ and using (3.5), (2.5) becomes

$$\Delta_j H_j \underline{w} = \lambda \Delta_j \underline{w}. \tag{3.7}$$

Since Δ_j is nonsingular, the result follows. $\square$

We now seek a similar formula for the roots of the residual polynomials. In the general case, the search is complicated by the possibility that the columns of S_j may not span $\mathcal{K}_j$. Consider the formulas (1.5). If $\alpha_i = 0$ for some i, then $\underline{s}_{i+1} = \underline{s}_i$ and the columns of S_j are not independent for $j > i$. However, since the columns of P_j do span $\mathcal{K}_j$, we can express S_j in terms of P_j. From (1.5e) and (1.5f) we have

$$\underline{s}_{i+1} - \underline{s}_i = -\alpha_i C A \underline{p}_i. \tag{3.8}$$

In matrix form this becomes

$$S_{j+1} \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 1 & -1 & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ & & 1 & -1 \\ & & & 1 \end{bmatrix}_{j+1 \times j} = -CAP_j \Lambda_j \tag{3.9}$$

where $\Lambda_j = diag(\alpha_0, \ldots, \alpha_{j-1})$. Using (3.3) we have

$$S_{j+1} \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 1 & -1 & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ & & 1 & -1 \\ & & & 1 \end{bmatrix}_{j+1 \times j} = -P_{j+1} \widetilde{H}_{j+1} \Lambda_j. \tag{3.10}$$

Now, from (1.5c) we see that $\underline{p}_0 = \underline{s}_0$. This extra equation can be combined with (3.4) and (3.10) to yield

$$S_{j+1} \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ & 1 & -1 & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ & & & 1 & -1 \\ & & & & 1 \end{bmatrix} = P_{j+1} \begin{bmatrix} 1 & \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0,j-1} \\ & 1 & \sigma_{11} & \cdots & \sigma_{1,j-1} \\ & & 1 & \ddots & \vdots \\ & & & \ddots & \sigma_{j-1,j-1} \\ & & & & 1 \end{bmatrix} D_{j+1},$$

where $D_{j+1} = diag(1, -\alpha_0, \ldots, -\alpha_{j-1})$. We denote this equation, with subscripts reduced by 1, as

$$S_j F_j = P_j U_j D_j. \tag{3.11}$$

Clearly, F_j and U_j are nonsingular. On the other hand, D_j will be singular whenever $\alpha_i = 0$ for some $i < j$. We offer the following result, which is an easy extension of Corollary 5 in [AMS90]:

Lemma 3.2. If BCA is definite, then $\alpha_i \neq 0$ for every i and, thus, D_j is nonsingular.

Proof. See Corollary 5 in [AMS90]. □

Now consider (3.3). From (1.5b) we have $\underline{p}_0 = \underline{s}_o = CA\underline{e}_0$. We may include this relationship in (3.3) by adding a new first column on each side to get

$$CA \begin{bmatrix} | & | & | & & | \\ \underline{e}_0 & \underline{p}_0 & \underline{p}_1 & \cdots & \underline{p}_{j-1} \\ | & | & | & & | \end{bmatrix} = P_{j+1} \begin{bmatrix} 1 & \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0,j-1} \\ & 1 & \sigma_{11} & \cdots & \sigma_{1,j-1} \\ & & 1 & \ddots & \vdots \\ & & & \ddots & \sigma_{j-1,j-1} \\ & & & & 1 \end{bmatrix},$$

which we denote, with reduced index, as

$$CA\tilde{P}_j = P_j U_j. \tag{3.12}$$

Letting $V_j = W_j = P_j$ in (2.29) we have

$$\lambda(P_j^* B(CA)^{-1} P_j)\underline{w} = (P_j^* B P_j)\underline{w} = \Delta_j \underline{w}. \tag{3.13}$$

Making use of (3.12) we have

$$\lambda(P_j^* B \tilde{P}_j U_j^{-1})\underline{w} = \Delta_j \underline{w}. \tag{3.14}$$

Letting $U_j \underline{v} = \underline{w}$ this becomes

$$\lambda(P_j^* B \tilde{P}_j)\underline{v} = \Delta_j U_j \underline{v}. \tag{3.15}$$

The first column of $P_j^* B \tilde{P}_j$ is given by

$$P_j^* B \underline{e}_0 = P_j^* B \sum_{i=0}^{n-1} \alpha_i \underline{p}_i = [\alpha_0 \delta_0, \ldots, \alpha_{j-1}\delta_{j-1}]^T. \tag{3.16}$$

Thus, we can write

$$P_j^* B \tilde{P}_j = \Delta_j \begin{bmatrix} \alpha_0 & 1 & 0 & \cdots & 0 \\ \alpha_1 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \ddots & 1 \\ \alpha_{j-1} & 0 & \cdots & \cdots & 0 \end{bmatrix}. \tag{3.17}$$

If we denote the last matrix as L_j, the roots we seek are the eigenvalues of the generalized eigenvalue problem

$$\lambda L_j \underline{v} = U_j \underline{v}. \tag{3.18}$$

A close look at L_j reveals that it is singular only if $\alpha_{j-1} = 0$, in which case $\underline{s}_j = \underline{s}_{j-1}$ and $\rho_j(\lambda) = \rho_{j-1}(\lambda)$. The roots of the residual polynomial are the same as on the previous step. If $\alpha_{j-1} \neq 0$, then

$$L_j^{-1} = \begin{bmatrix} 0 & \cdots & \cdots & 0 & \frac{1}{\alpha_{j-1}} \\ 1 & \ddots & & \vdots & -\frac{\alpha_0}{\alpha_{j-1}} \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -\frac{\alpha_{j-2}}{\alpha_{j-1}} \end{bmatrix}. \tag{3.19}$$

Finally, the roots of the residual polynomial are the eigenvalues of

$$U_j L_j^{-1} = \begin{bmatrix} \widetilde{H}_{j-1} & \vdots & \underline{\eta}_{j-1} \end{bmatrix} = \begin{bmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0,j-2} & \eta_{0,j-1} \\ 1 & \sigma_{11} & \cdots & \sigma_{1,j-2} & \eta_{1,j-1} \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{j-2,j-2} & \eta_{j-2,j-1} \\ 0 & \cdots & 0 & 1 & \eta_{j-1,j-1} \end{bmatrix}, \tag{3.20}$$

where $\widetilde{H}_{j-1}$ is defined in (3.4) and

$$\underline{\eta}_{j-1} = \begin{bmatrix} \eta_{0,j-1} \\ \eta_{1,j-1} \\ \vdots \\ \eta_{j-1,j-1} \end{bmatrix} = \frac{1}{\alpha_{j-1}} U_j \begin{bmatrix} 1 \\ -\alpha_0 \\ \vdots \\ -\alpha_{j-2} \end{bmatrix}. \tag{3.21}$$

If we compare H_j (3.6) and $U_j L_j^{-1}$ (3.20) we see they differ only in the last column. In fact, the vector $\underline{\eta}_j$ can be computed as soon as α_{j-1} is computed, that is, before the last column of H_j is available. It can be accumulated recursively as follows. Let

$$\underline{\xi}_{i-1} = U_i \begin{bmatrix} 1 \\ -\alpha_0 \\ \vdots \\ -\alpha_{i-2} \end{bmatrix}, \tag{3.22}$$

$$\underline{\xi}_i = \begin{bmatrix} | \\ \underline{\xi}_{i-1} \\ | \\ 0 \end{bmatrix} - \alpha_{i-1} \begin{bmatrix} \sigma_{0,i-1} \\ \vdots \\ \sigma_{i-1,i-1} \\ 1 \end{bmatrix}. \tag{3.23}$$

If $\alpha_i \neq 0$, then

$$\underline{\eta}_i = \frac{1}{\alpha_i} \underline{\xi}_i. \tag{3.24}$$

Note that (3.21) yields

$$\eta_{i,i} = -\frac{\alpha_{i-1}}{\alpha_i}. \tag{3.25}$$

We summarize the above discussion in the following Theorem:

Theorem 3.3. The roots of the residual polynomial, $\rho_j(\lambda) = 0$, are the eigenvalues of the generalized eigenvalue problem

$$\lambda L_j \underline{v} = U_j \underline{v}. \tag{3.26}$$

If $\alpha_{j-1} = 0$, then $\rho_j(\lambda) = \rho_{j-1}(\lambda)$. If $\alpha_{j-1} \neq 0$, then the roots are the eigenvalues of

$$U_j L_j^{-1} = \begin{bmatrix} \widetilde{H}_{j-1} & \vdots & \underline{\eta}_{j-1} \end{bmatrix}. \tag{3.27}$$

Proof. The proof follows from the discussion above. □

3.2. The Orthomin Algorithm

If BCA is definite, then $\alpha_i \underline{p}_i \neq \underline{0}\ \forall\ i$ and the columns of S_i span the space $\mathcal{K}_i(\underline{s}_0, CA)$. Thus, one may replace $CA\underline{p}_i$ by $\underline{s}_{i+1}$ in (1.5g). The result is the Orthomin algorithm. The Orthomin algorithm is less expensive and somewhat more stable than the Orthodir algorithm [YoJe80], [JeYo83]. In this section we will show how the roots of the orthogonal polynomials and the residual polynomials can be computed from information computed using the Orthomin algorithm. As in (1.5), given $\underline{x}_0$ let

$$
\begin{aligned}
&\underline{r}_0 = b - A\underline{x}_0, && && (3.28a)\\
&\underline{s}_0 = C\underline{r}_0, && && (3.28b)\\
&\underline{\hat{p}}_0 = \underline{s}_0, && && (3.28c)\\
&\vdots && && \\
&\underline{x}_{j+1} = \underline{x}_j + \hat{\alpha}_j \underline{\hat{p}}_j, && \hat{\alpha}_j = \frac{\langle B\underline{e}_j, \underline{\hat{p}}_j\rangle}{\langle B\underline{\hat{p}}_j, \underline{\hat{p}}_j\rangle} = \frac{\langle B\underline{e}_j, \underline{s}_j\rangle}{\langle B\underline{\hat{p}}_j, \underline{\hat{p}}_j\rangle}, && (3.28d)\\
&\underline{r}_{j+1} = \underline{r}_j - \hat{\alpha}_j A\underline{\hat{p}}_j, && && (3.28e)\\
&\underline{s}_{i+1} = C\underline{r}_{i+1}, && && (3.28f)\\
&\underline{\hat{p}}_{j+1} = \underline{s}_{j+1} - \textstyle\sum_{i=0}^{j} \hat{\sigma}_{ij}\underline{\hat{p}}_i, && \hat{\sigma}_{ij} = \frac{\langle B\underline{s}_{j+1}, \underline{\hat{p}}_i\rangle}{\langle B\underline{\hat{p}}_i, \underline{\hat{p}}_i\rangle}. && (3.28g)
\end{aligned}
$$

Since the sequence $\{\underline{x}_i\}$ is uniquely determined by (1.2) and (1.3) we see from (1.5d) and (3.28d) that

$$\alpha_j \underline{p}_j = \hat{\alpha}_j \underline{\hat{p}}_j. \tag{3.29}$$

Since BCA was assumed to be definite, Lemma 3.2 yields $\hat{\alpha}_j \neq 0$. Following arguments in [AMS90] Section 4 we can show that

$$\hat{\alpha}_j = -\frac{\alpha_j}{\alpha_{j-1}} \tag{3.30}$$

and thus,

$$\underline{\hat{p}}_j = -\alpha_{j-1}\underline{p}_j. \tag{3.31}$$

From (1.5c) and (3.28c) we have $\underline{\hat{p}}_0 = \underline{p}_0 = \underline{s}_0$ which yields

$$\widehat{P}_j = P_j D_j, \tag{3.32}$$

where $D_j = diag(1, -\alpha_0, \ldots, -\alpha_{j-2})$ is defined in (3.11). As in (3.2), we know from (1.6) that

$$\widehat{P}_j^* B \widehat{P}_j = \widehat{\Delta}_j = diag(\ldots, \hat{\delta}_i, \ldots), \tag{3.33}$$

where $\hat{\delta}_i = \langle B\underline{\hat{p}}_i, \underline{\hat{p}}_i\rangle$.

Consider the recurrence (3.28g). In matrix form this becomes

$$S_j = \widehat{P}_j \begin{bmatrix} 1 & \hat{\sigma}_{00} & \hat{\sigma}_{01} & \cdots & \hat{\sigma}_{0,j-2} \\ & 1 & \hat{\sigma}_{11} & \cdots & \hat{\sigma}_{1,j-2} \\ & & 1 & \ddots & \vdots \\ & & & \ddots & \hat{\sigma}_{j-2,j-2} \\ & & & & 1 \end{bmatrix} \equiv \widehat{P}_j \widehat{U}_j. \tag{3.34}$$

Together with (3.32) and (3.11) we have

$$\widehat{U}_j F_j = D_j^{-1} U_j D_j. \tag{3.35}$$

As in (3.9), the relations (3.28e) and (3.28f) yield

$$S_{j+1} \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 1 & -1 & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ & & 1 & -1 \\ & & & 1 \end{bmatrix}_{j+1 \times j} = -CA\widehat{P}_j \widehat{\Lambda}_j \tag{3.36}$$

where $\widehat{\Lambda}_j = diag(\hat{\alpha}_0, \cdots, \hat{\alpha}_{j-1})$. Rearranging (3.36) yields

$$S_j \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}_{j \times j} = S_j F_j^* = CA\widehat{P}_j \widehat{\Lambda}_j + \underline{s}_j \underline{c}_j^*, \tag{3.37}$$

where $\underline{c}_j = [0, \ldots, 0, 1]^*$ is the jth canonical basis element.

We first find the roots of the residual polynomials, $\rho(\lambda) = 0$. Let $W_j = \widehat{P}_j$ and $V_j = S_j F_j^*$ in (2.29) to get

$$\lambda(\widehat{P}_j^* B(CA)^{-1} S_j F_j^*)\underline{w} = (\widehat{P}_j^* B S_j F_j^*)\underline{w}. \tag{3.38}$$

Using (3.37) and (3.33) the left-hand side becomes

$$\widehat{P}_j^* B \widehat{P}_j \widehat{\Lambda}_j + \widehat{P}_j^* B(CA)^{-1} \underline{s}_j \underline{c}_j^* = \widehat{\Delta}_j \widehat{\Lambda}_j, \tag{3.39}$$

the second term being zero by (1.7). Using (3.34) and (3.33) the right-hand side becomes

$$\widehat{P}_j^* B S_j F_j^* = \widehat{P}_j^* B \widehat{P}_j \widehat{U}_j F_j^* = \widehat{\Delta}_j \widehat{U}_j F_j^*. \tag{3.40}$$

Thus, the roots we seek are the eigenvalues of $\widehat{U}_j F_j^* \widehat{\Lambda}_j^{-1}$.

To find the roots of the orthogonal polynomial, $\phi_j(\lambda) = 0$, let $W_j = \widehat{P}_j$ and $V_j = \widehat{P}_j \widehat{\Lambda}_j$ in (2.5) to get

$$(\widehat{P}_j^* B(CA)\widehat{P}_j \widehat{\Lambda}_j)\underline{w} = \lambda(\widehat{P}_j^* B \widehat{P}_j \widehat{\Lambda}_j)\underline{w} = \lambda(\widehat{\Delta}_j \widehat{\Lambda}_j)\underline{w} \tag{3.41}$$

Using (3.37) and (3.34) the left-hand side becomes

$$\widehat{P}_j^* B(CA)\widehat{P}_j \widehat{\Lambda}_j = \widehat{P}_j^* B \widehat{P}_j \widehat{U}_j F_j^* - \widehat{P}_j^* B \underline{s}_j \underline{c}_j^* = \widehat{\Delta}_j \widehat{U}_j F_j^* - \widehat{P}_j^* B \underline{s}_j \underline{c}_j^*. \tag{3.42}$$

From (3.28g) we have that

$$\widehat{\Delta}_j^{-1}\widehat{P}_j^* B\underline{s}_j = \underline{\hat{\sigma}}_{j-1} = \begin{bmatrix} \hat{\sigma}_{0,j-1} \\ \hat{\sigma}_{1,j-1} \\ \vdots \\ \hat{\sigma}_{j-1,j-1} \end{bmatrix}. \tag{3.43}$$

Finally, the roots we seek are the eigenvalues of

$$\widehat{U}_j F_j^* \widehat{\Lambda}_j^{-1} - \frac{1}{\hat{\alpha}_{j-1}}\underline{\hat{\sigma}}_{j-1}\underline{c}_j^*. \tag{3.44}$$

As in the Orthodir implementation the two eigenvalue problems differ by a rank-one matrix.

We summarize the above discussion in the following Theorem.

Theorem 3.4. Using the Orthomin algorithm, the roots of the residual polynomial, ρ_j are the eigenvalues of

$$\widehat{U}_j F_j^* \widehat{\Lambda}_j^{-1} = \begin{bmatrix} 1 & \hat{\sigma}_{00} & \cdots & \hat{\sigma}_{0,j-2} \\ & 1 & \ddots & \vdots \\ & & \ddots & \hat{\sigma}_{j-2,j-2} \\ & & & 1 \end{bmatrix}_{j\times j} \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}_{j\times j} \widehat{\Lambda}_j^{-1}.$$

The roots of the orthogonal polynomial, ϕ_j, are the eigenvalues of

$$\widehat{U}_j F_j^* \widehat{\Lambda}_j^{-1} - \frac{1}{\hat{\alpha}_{j-1}}\underline{\hat{\sigma}}_{j-1}\underline{c}_j^* =$$

$$\begin{bmatrix} 1 & \hat{\sigma}_{00} & \cdots & \hat{\sigma}_{0,j-2} & \hat{\sigma}_{0,j-1} \\ & 1 & \ddots & \vdots & \vdots \\ & & \ddots & \hat{\sigma}_{j-2,j-2} & \hat{\sigma}_{j-2,j-1} \\ & & & 1 & \hat{\sigma}_{j-1,j-1} \end{bmatrix}_{j\times j+1} \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ & & & -1 \end{bmatrix}_{j+1\times j} \widehat{\Lambda}_j^{-1}.$$

Proof. The proof follows from the discussion above. □

3.3. The Special Case CA B-Self-Adjoint

The above discussion simplifies somewhat when a three term recursion exists. First let us consider the case when CA is B-self-adjoint, that is, when

$$\langle BCA\underline{x}, \underline{y}\rangle = \langle B\underline{x}, CA\underline{y}\rangle, \tag{3.45}$$

for every $\underline{x}, \underline{y} \in \mathcal{C}^n$. Equivalently,

$$\begin{aligned} BCA &= (CA)^* B = (BCA)^*, && (3.46) \\ B(CA)^{-1} &= (CA)^{-*} B = (B(CA)^{-1})^*. && (3.47) \end{aligned}$$

From (3.46) and (3.5) we see that $P_j^* BCAP_j = \Delta_j H_j$ is Hermitian, which implies that

$$\begin{aligned} \sigma_{i,i} &\in \mathcal{R} && (3.48) \\ \sigma_{i,j} &= 0 \qquad \text{for} \qquad j > i+1 && (3.49) \\ \sigma_{i,i+1} &= \frac{\delta_i}{\delta_{i+1}}. && (3.50) \end{aligned}$$

This, in turn, implies that H_j is similar to

$$\Delta_j^{1/2} H_j \Delta_j^{-1/2} = \begin{bmatrix} \sigma_{00} & \sqrt{\sigma_{01}} & & \\ \sqrt{\sigma_{01}} & \sigma_{11} & \ddots & \\ & \ddots & \ddots & \sqrt{\sigma_{j-2,j-1}} \\ & & \sqrt{\sigma_{j-2,j-1}} & \sigma_{j-1,j-1} \end{bmatrix}_{j \times j}, \qquad (3.51)$$

which is a real symmetric tridiagonal matrix. Thus, the eigenvalues of the H_j are real, are separated by the eigenvalues of H_{j-1} and converge to the eigenvalues of CA (cf. [SlVo86]). Moreover, from Lemma 2.5 we see that $\mathcal{H}(H_j) \subseteq \mathcal{H}(CA) \subset \mathcal{R}$.

Likewise, from (3.47), (3.12) and (3.17) we see that

$$P_j^* B(CA)^{-1} P_j = \Delta_j L_j U_j^{-1} \qquad (3.52)$$

is Hermitian, which implies that $\Delta_j^{1/2} L_j U_j^{-1} \Delta_j^{-1/2}$ is also Hermitian, which, in turn, implies that the eigenvalues of (3.18) are real. Moreover, if $\alpha_{j-1} \neq 0$ then $U_j L_j^{-1} \Delta_j^{-1}$ exists and is Hermitian. In addition to (3.48),(3.49) and (3.50) this implies

$$\begin{aligned} \eta_{i,i} &\in \mathcal{R} && (3.53) \\ \eta_{i,j} &= 0 \qquad \text{for} \qquad j > i+1 && (3.54) \\ \eta_{i,i+1} &= \frac{\delta_i}{\delta_{i+1}}. && (3.55) \end{aligned}$$

Thus, in this case, $U_j L_j^{-1}$ differs from H_j in only the bottom right corner. We have

$$U_j L_j^{-1} = H_j + (\eta_{j-1,j-1} - \sigma_{j-1,j-1}) \underline{e}_j \underline{e}_j^*, \qquad (3.56)$$

where, again, $\underline{e}_j = [0, \ldots, 0, 1]^*$. This leads to the following theorem which summarizes the above discussion.

Theorem 3.5. Let CA be B-self-adjoint and $\alpha_{j-1} \neq 0$. Then, the roots of $\phi_j(\lambda) = 0$ are real, are separated by the roots of $\phi_{j-1}(\lambda) = 0$ and converge to eigenvalues of CA. The roots of $\rho_j(\lambda) = 0$ are real, are separated by the roots of $\phi_{j-1}(\lambda) = 0$ and converge to eigenvalues of CA. Further, let

$$\lambda_1^k < \lambda_2^k \cdots < \lambda_k^k \tag{3.57}$$

be the roots of $\phi_k(\lambda) = 0$ and let

$$\hat{\lambda}_1^k < \hat{\lambda}_2^k \cdots < \hat{\lambda}_k^k \tag{3.58}$$

be the roots of $\rho_k(\lambda) = 0$. If $(\eta_{j-1,j-1} - \sigma_{j-1,j-1}) > 0$, then

$$\lambda_1^j < \hat{\lambda}_1^j < \lambda_1^{j-1} < \lambda_2^j < \cdots < \hat{\lambda}_{j-1}^j < \lambda_{j-1}^{j-1} < \lambda_j^j < \hat{\lambda}_j^j. \tag{3.59}$$

If $(\eta_{j-1,j-1} - \sigma_{j-1,j-1}) < 0$, then

$$\hat{\lambda}_1^j < \lambda_1^j < \lambda_1^{j-1} < \hat{\lambda}_2^j < \cdots < \lambda_{j-1}^j < \lambda_{j-1}^{j-1} < \hat{\lambda}_j^j < \lambda_j^j. \tag{3.60}$$

Proof. These results follow immediately from well know results on the eigenvalues of Hermitian tridiagonal matrices (cf. [Go73], [GoVL89]). We offer the following brief proof for completeness. Consider the eigenvalues of the Hermitian matrices $\Delta_j^{1/2} H_j \Delta_j^{-1/2}$ and $\Delta_j^{1/2} U_j L_j^{-1} \Delta_j^{-1/2}$, which are similar to H_j and $U_j L_j^{-1}$ respectively. Since $\Delta_{j-1}^{1/2} H_{j-1} \Delta_{j-1}^{-1/2}$ is a principal submatrix of both, its eigenvalues separate the eigenvalues of both matrices. Now, from (3.56) we have

$$\Delta_j^{1/2} U_j L_j^{-1} \Delta_j^{-1/2} = \Delta_j^{1/2} H_j \Delta_j^{-1/2} + (\eta_{j-1,j-1} - \sigma_{j-1,j-1}) \underline{e}_j \underline{e}_j^*. \tag{3.61}$$

Using the minimax definition of the eigenvalues of a Hermitian matrix (cf. [St73]) we see that

$$(\eta_{j-1,j-1} - \sigma_{j-1,j-1}) > 0 \quad \Rightarrow \quad \lambda_i^j < \hat{\lambda}_i^j, \tag{3.62}$$

$$(\eta_{j-1,j-1} - \sigma_{j-1,j-1}) < 0 \quad \Rightarrow \quad \lambda_i^j > \hat{\lambda}_i^j, \tag{3.63}$$

for $i = 1, \ldots, j$. The separation property of the eigenvalues of H_{j-1} completes the proof. □

Corollary 3.6. Suppose CA is B-self-adjoint indefinite with $\Sigma(CA) \subseteq [a, b] \cup [c, d]$ where $a, b, c, d \in \Sigma(CA)$ and $b < 0 < c$. Then, at most one root of $\phi_j(\lambda) = 0$ can lie in (b, c) while at most one root of $\rho_j(\lambda) = 0$ may lie outside

$[a,d]$. If $(\eta_{j-1,j-1} - \sigma_{j-1,j-1}) > 0$ it may lie to the right of $[a,d]$, while if $(\eta_{j-1,j-1} - \sigma_{j-1,j-1}) < 0$ it may lie to the left of $[a,d]$.

Proof. By Corollary 2.2 no root of $\phi_j(\lambda) = 0$ can lie outside $[a,d]$ and by Corollary 2.4 no root of $\rho_j(\lambda) = 0$ can lie inside (b,c). From Theorem 3.5 the roots of $\rho_j(\lambda) = 0$ interleave the roots of $\phi_j(\lambda) = 0$, shifted according to the sign of $(\eta_{j-1,j-1} - \sigma_{j-1,j-1})$. $\square$

3.4. The Special Case CA B-Normal(1)

In this section we consider the case in which CA is B-normal(1) but not B-self-adjoint. As stated in (2.36), we may write

$$CA = e^{i\theta}(G + irI) \tag{3.64}$$

where $r \in \mathcal{R}$ and G is B-self-adjoint. In this section we will assume that $r \neq 0$. The case in which $r = 0$ but $\theta \neq 0$ is a simple scaling of the case in which CA is B-self-adjoint. Recall from Section 2 that the eigenvalues of G lie in a real interval $[a,b]$ and thus, $\Sigma(CA)$ is contained in the line segment $[e^{i\theta}(a+ir), e^{i\theta}(b+ir)]$.

We first consider the roots of the orthogonal polynomials, ϕ_j. From (3.64) we see that

$$P_j^* BCAP_j = e^{i\theta}(P_j^* BGP_j + ir\Delta_j) = \Delta_j H_j. \tag{3.65}$$

Since $P_j^* BGP_j$ is Hermitian, $\Delta_j H_j$ is a translation and scaling of a Hermitian matrix. Let

$$T_j = \Delta_j^{-1/2} P_j^* BGP_j \Delta_j^{-1/2}. \tag{3.66}$$

Clearly, T_j is a tridiagonal Hermitian matrix and an orthogonal section of G. As such its eigenvalues are contained in the real interval $[a,b]$ and are separated by the eigenvalues of T_{j-1}. The eigenvalues of H_j are the eigenvalues of $\Delta_j^{1/2} H_j \Delta_j^{-1/2}$ which are a scaling and translation of the eigenvalues of T_j. Thus, the eigenvalues of H_j lie on the line segment $[e^{i\theta}(a+ir), e^{i\theta}(b+ir)]$ and are separated along this line segment by the eigenvalues of H_{j-1}. For a more detailed presentation of this property see [AMS90], Section 7.

Now consider the roots of the residual polynomials, ρ_j. Since we have assumed that $r \neq 0$ we know that BCA is definite and thus, $\alpha_i \neq 0$ for all i. The columns of S_j span $\mathcal{K}_j$. We may use a variant of (3.9) to get

$$S_j \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}_{j \times j} = S_j F_j^* = CAP_j\Lambda_j + \underline{s}_j \underline{c}_j^*, \tag{3.67}$$

which is similar to (3.37). Letting $V_j = S_j F_j^*$ and $W_j = P_j$ in (2.29) we have the eigenvalue problem

$$\lambda(P_j^* B(CA)^{-1} S_j F_j^*)\underline{w} = (P_j^* B S_j F_j^*)\underline{w}. \tag{3.68}$$

Making use of (3.67) we have

$$P_j^* B(CA)^{-1} S_j F_j^* = P_j^* B P_j \Lambda_j + P_j^* B(CA)^{-1} \underline{s}_j \underline{c}_j = \Delta_j \Lambda_j, \tag{3.69}$$

the rank-one term being zero by (1.7). Similarly,

$$P_j^* B S_j F_j^* = P_j^* B(CA) P_j + P_j^* B \underline{s}_j \underline{c}_j = \Delta_j H_j \Lambda_j + P_j^* B \underline{s}_j \underline{c}_j. \tag{3.70}$$

Putting (3.69) and (3.70) together we conclude that the roots of $\rho_j(\lambda) = 0$ are the eigenvalues of

$$\lambda \underline{w} = (H_j + \frac{1}{\alpha_{j-1}} \Delta_j^{-1} P_j^* B \underline{s}_j \underline{c}_j^*)\underline{w}. \tag{3.71}$$

Comparing this derivation with the derivation in Section 3.1 yields

$$U_j L_j^{-1} = H_j + \frac{1}{\alpha_{j-1}} \Delta_j^{-1} P_j^* B \underline{s}_j \underline{c}_j^*. \tag{3.72}$$

Now, suppose that CA is B-normal(1). From (3.64) we have

$$(CA)^\dagger = e^{-2i\theta} CA - 2ie^{-i\theta} rI, \tag{3.73}$$

which yields

$$\langle B\underline{s}_j, \underline{p}_k \rangle = \langle BCA\underline{e}_j, \underline{p}_k \rangle = e^{2i\theta} \langle B\underline{e}_j, CA\underline{p}_k \rangle + 2ie^{i\theta} \langle B\underline{e}_j, \underline{p}_k \rangle. \tag{3.74}$$

Using (1.7) we have

$$\langle B\underline{s}_j, \underline{p}_k \rangle = \left\{ \begin{array}{cc} 0 & k < j-1 \\ e^{2i\theta} \langle B\underline{e}_j, \underline{p}_j \rangle & k = j-1 \end{array} \right. . \tag{3.75}$$

This yields

$$U_j L_j^{-1} = H_j + e^{2i\theta} \frac{\langle B\underline{e}_j, \underline{p}_j \rangle}{\langle B\underline{e}_{j-1}, \underline{p}_{j-1} \rangle} \underline{c}_j \underline{c}_j^*. \tag{3.76}$$

Again, the roots of the residual polynomial can be found as the eigenvalues of a rank-one update to H_j. Letting

$$\gamma_{j-1} = e^{i\theta} \frac{\langle B\underline{e}_j, \underline{p}_j \rangle}{\langle B\underline{e}_{j-1}, \underline{p}_{j-1} \rangle} \tag{3.77}$$

and using (3.65) and (3.66) we may write

$$\Delta^{1/2} U_j L_j^{-1} \Delta^{-1/2} = e^{i\theta}((T_j + \gamma_{j-1} \underline{c}_j \underline{c}_j^*) + irI), \tag{3.78}$$

and the eigenvalues we seek are a scaling and translation of the eigenvalues of $T_j + \gamma_{j-1}\underline{c}_j\underline{c}_j^*$. Unfortunately, γ_{j-1} need not be real, and thus, the eigenvalues of $T_j + \gamma_{j-1}\underline{c}_j\underline{c}_j^*$ may not be real. For more details on the eigenvalues of a general tridiagonal matrix perturbed by a rank-one update see [AGG88] and [ArGo88]. Of course, the roots of the residual polynomial will be contained in the set $\mathcal{F}_B^{-1}((CA)^{-1})$ as described in Section 2.3.

We close this section by noting that when CA is B-normal(1) the Orthomin algorithm is also greatly simplified. Most notably,

$$\begin{aligned} \hat{\sigma}_{i,j} &= 0 \quad \text{for} \quad j > i, && (3.79) \\ \hat{\sigma}_{i,i} &= -e^{2i\theta} \frac{\langle B\underline{e}_{i+1}, \underline{s}_{i+1}\rangle}{\langle B\underline{e}_i, \underline{s}_i\rangle} = -e^{2i\theta} \frac{\langle B\underline{e}_{i+1}, \underline{\hat{p}}_{i+1}\rangle}{\langle B\underline{e}_i, \underline{\hat{p}}_i\rangle}. && (3.80) \end{aligned}$$

(See [AMS90] Section 4 for details.) Thus, $\widehat{U}_j$ is bidiagonal. A glance at the equations in Theorem 3.4 yields the observation that the products of the matrices are tridiagonal and that the rank-one change effects only the lower right corner.

4. Approximating $\boldsymbol{\Sigma(CA)}$

In this Section we will present an algorithm for constructing a set that describes the spectrum of CA from information gathered during a conjugate gradient iteration. Such a set could be used to construct a polynomial preconditioning or a polynomial iterative method. A desirable property of such a set is that it does not contain the origin.

Most previous work has proposed methods in which a set of eigenvalue estimates is obtained in some fashion and the convex hull of the estimates is used to approximate $\Sigma(CA)$ (cf. [Ma78], [As85]). Unfortunately, the best one can expect from a set of estimates is that they are contained in $\mathcal{F}_M(CA)$ for some M. If MCA is indefinite, that is, $0 \in \mathcal{F}_M(CA)$, then the hull of the estimates may also contain the origin.

This problem was avoided in [NRT92]. There the roots of the residual polynomial of a fixed number of steps of a conjugate gradient iteration were used to construct a Richardson type polynomial iteration. A level set of the residual polynomial was used to describe the pseudo-spectrum of CA. There are two problems with this approach. First, as we will see in the next section, the roots of the residual polynomial can sometimes lie significantly outside of $\mathcal{F}_B(CA)$. A second problem is that it is difficult to determine which level set yields the best region. Of course, this does not impact the iteration, only the determination of a set which describes the pseudo-spectrum.

Here we make use of the observation that

$$\Sigma(CA) \subseteq \mathcal{G} \equiv \mathcal{F}_B(CA) \cap \mathcal{F}_B^{-1}((CA)^{-1}). \tag{4.1}$$

In fact, $\mathcal{G}$ is contained in an annulus:

$$\mathcal{G} \subseteq \{z : \frac{1}{\|(CA)^{-1}\|_B} \leq z \leq \|CA\|_B\}. \tag{4.2}$$

From Corollary 2.2 we know that the roots of the orthogonal polynomials are contained in $\mathcal{F}_B(CA)$, that is,

$$\mathcal{H}_j \equiv \mathcal{H}(\{\lambda_i^j\}_{i=1}^j) \subseteq \mathcal{F}_B(CA). \tag{4.3}$$

Likewise, from Corollary 2.4 we know that the reciprocals of the roots of the residual polynomials are contained in $\mathcal{F}_B((CA)^{-1})$, that is,

$$\widehat{\mathcal{H}}_j \equiv \mathcal{H}(\{\frac{1}{\hat{\lambda}_i^j}\}_{i=1}^j) \subseteq \mathcal{F}_B((CA)^{-1}). \tag{4.4}$$

Let us denote

$$\widehat{\mathcal{H}}_j^{-1} \equiv \{z : \frac{1}{z} \in \mathcal{H}(\{\frac{1}{\hat{\lambda}_i^j}\}_{i=1}^j)\}. \tag{4.5}$$

Clearly,

$$\mathcal{H}_j \cap \widehat{\mathcal{H}}_j^{-1} \subseteq \mathcal{G}. \tag{4.6}$$

This yields the following algorithm:

1. Given the linear system (1.1) with preconditioning C, inner product matrix B and initial guess $\underline{x}_0$, begin either the Orthodir or Orthomin iteration whichever is applicable.

2. At each step, or after a predetermined number of steps, compute $\{\lambda_i^j\}_{i=1}^j$ and $\{\hat{\lambda}_i^j\}_{i=1}^j$, the roots of the orthogonal polynomials and the residual polynomials respectively.

3. Form the convex polygons

$$\begin{aligned} \mathcal{H}_j &\equiv \mathcal{H}(\{\lambda_i^j\}_{i=1}^j) \\ \widehat{\mathcal{H}}_j &\equiv \mathcal{H}(\{\frac{1}{\hat{\lambda}_i^j}\}_{i=1}^j). \end{aligned}$$

4. Form the set

$$\mathcal{G}_j = \mathcal{H}_j \cap \widehat{\mathcal{H}}_j^{-1}.$$

Before proceeding to numerical examples we make some remarks. First, notice that the convex polygons $\mathcal{H}_j$ and $\widehat{\mathcal{H}}_j$ can be formed by taking the convex hull of the union of all previous sets. For example, if the roots are computed on every step of the conjugate gradient iteration we might define

$$\begin{aligned} \mathcal{H}_j &\equiv \mathcal{H}(\cup_{k=1}^{j}\{\lambda_i^k\}_{i=1}^k) \\ \widehat{\mathcal{H}}_j &\equiv \mathcal{H}(\cup_{k=1}^{j}\{\frac{1}{\hat{\lambda}_i^k}\}_{i=1}^k). \end{aligned}$$

We also note that while $\mathcal{H}_j$ and $\widehat{\mathcal{H}}_j$ are convex polygons, the set $\widehat{\mathcal{H}}_j^{-1}$ will not have straight sides. Its appearance can be quite interesting. It is the image of a polygon under a Moebius transformation. Thus, the boundary consists of circular arcs and intervals. Finally, as we will see in the next section, the set $\mathcal{G}_j$ may be empty.

5. Numerical Results

In this final section we report some numerical results. First, we address the B-self-adjoint indefinite case, then the B-normal(1) case and finally three examples of the general case. In the general case, for well-conditioned definite systems, the set $\mathcal{G}_j$ yields good information for small values of j. As the matrix becomes ill-conditioned, or as the matrix becomes indefinite, more iterations are required before $\mathcal{G}_j$ yields useful information. The examples we have chosen are near the boundary of the applicability of this technique.

5.1. *B*-Self-Adjoint Indefinite

In this example a diagonal matrix was constructed with 40 eigenvalues uniformly distributed in two intervals (20 each), specifically, $\Sigma(A) \subseteq [-2,-1] \cup [.5, 1.5]$. The Orthodir algorithm was used to implement a conjugate gradient iteration with $C = I$ and $B = A^*A$. Figure 2.a shows the results of a typical step. Notice that one root of the residual polynomial lies to the right of d while one root of the orthogonal polynomial is in the interval (b, c). Also, notice that the roots of the two polynomials are interleaved. Figure 2.b is included to demonstrate that a root of the residual polynomial may lie well outside $\mathcal{H}(A)$ even after a relatively large number of steps.

5.2. B-Normal(1)

In this example a diagonal matrix was constructed with 40 eigenvalues uniformly distributed on the line segment $[e^{-.5i}(-1+i), e^{-.5i}(2+i)]$. Again, an Orthodir algorithm was used to implement a conjugate gradient iteration with $C = I$ and $B = A^*A$. The initial residual was chosen randomly. Figure 3.a illustrates that the first root of the residual polynomial, denoted with a $+$, may lie anywhere in $\mathcal{F}_B^{-1}(A^{-1})$. In fact, if the initial residual were chosen to be a linear combination of the two eigenvectors associated with the largest and smallest eigenvalues, then the first root would lie on the boundary of $\mathcal{F}_B^{-1}(A^{-1})$. As expected, the roots of the orthogonal polynomials, denoted with an $*$, lie on the line segment. As the iteration progresses the roots of the residual polynomials will begin to converge to the line segment, never quite reaching it for all values until the final step. This is exemplified in Figure 3.b.

5.3. Annulus Segment

In this experiment the matrix A was constructed by choosing a uniformly distributed set of values $\{r_j\}_{j=1}^m \subset [1-\zeta, 1+\zeta]$ and a uniformly distributed set of values $\{\theta_k\}_{k=1}^{\ell} \subset [-\theta, \theta]$. A real $m\ell \times m\ell$ matrix was then constructed to have the eigenvalues $\lambda_{j,k} = r_j e^{i\theta_k}$. The eigenvalues lie in a segment of an annulus in the complex plane (see Figure 4.a and Figure 5.a). The matrix was constructed by first forming a block diagonal matrix with 2×2 blocks whose eigenvalues were a conjugate pair. Then, a randomly generated similarity transformation was applied to insure non-normality.

Figure 4.a is a plot of the eigenvalues with $\zeta = .4$, $\theta = 1.5$, $m = 6$ and $\ell = 15$. The curves plotted in that figure are the $\mathcal{H}(A)$ and $\mathcal{H}^{-1}(A^{-1})$. Because θ is so close to $\pi/2$, $\mathcal{H}^{-1}(A^{-1})$ includes most of the right-half plane. The set $\mathcal{H}(A) \cap \mathcal{H}^{-1}(A^{-1})$ fits the spectrum snugly. A more illustrative example is shown in Figure 5.a, which uses the values $\zeta = .4$, $\theta = 2.0$, $m = 6$ and $\ell = 15$. Here, the set $\mathcal{H}^{-1}(A^{-1})$ is the area outside the curve that passes through the inner-most ring of eigenvalues. This is because $\mathcal{H}(A^{-1})$ contains the origin.

The conjugate gradient iteration was performed with $C = I$ and $B = A^*A$. Figures 4.b-4.d and Figures 5.b-5.d show the roots of the orthogonal polynomial, denoted by $*$, and their hull, $\mathcal{H}_j$, and the roots of the residual polynomial, denoted by $+$, and their reciprocal hull, $\widehat{\mathcal{H}}_j^{-1}$, for various values of j. Since the matrices in this example are not B-normal, the theory only guarantees that $\mathcal{G}_j \subseteq \mathcal{F}_B(A) \cap \mathcal{F}_B^{-1}((CA)^{-1})$. However, after an appropriate number of iterations the set $\mathcal{G}_j$ begins to approximate $\mathcal{G}_n = \mathcal{H}(A) \cap \mathcal{H}^{-1}(A^{-1})$. The appropriate number seemed to depend upon ζ and θ. For $\theta \ll \pi/2$ few steps were required. As θ grew more steps were required.

Notice that in Figure 5.b, the set $\mathcal{G}_j$ is empty. This was typical behavior in the first few iterations. The roots of the orthogonal polynomials seemed to be oblivious to the origin while the roots of the residual polynomials circled their prey cautiously.

Round-off error did not seem to play a role in these experiments. When the iteration was carried out the full $n = m\ell$ steps, the roots and eigenvalues lined up to within visual limits. In Figure 5.d many of the roots have already lined up. This was true in all of our experiments. We also remark that saving all previous roots to form the sets $\mathcal{H}_j$ and $\widehat{\mathcal{H}}_j^{-1}$ did speed up the process but not significantly.

5.4. The Grcar Matrix

This example involved the 104×104 Toeplitz matrix:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 1 & 1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 1 & 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 0 & -1 & 1 & 1 & 1 \\ 0 & 0 & \cdots & 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & \cdots & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

A conjugate gradient method was implemeted with $C = I$ and $B = A^*A$. Figure 6.a shows the eigenvalues of A and the sets $\mathcal{H}(A)$ and $\mathcal{H}^{-1}(A^{-1})$. The boundary of the set $\mathcal{H}^{-1}(A^{-1})$ runs from the bottom edge of the figure, hugs the left side of the spectrum and then exits the figure near the top. It includes a large area in the right-half plane. The set $\mathcal{H}(A)$ is more easily seen. Figures 6.b-6.d again show the roots of the orthogonal polynomial, denoted by $*$, and their hull, $\mathcal{H}_j$, and the roots of the residual polynomial, denoted by $+$, and their reciprocal hull, $\widehat{\mathcal{H}}_j^{-1}$, for various values of j. In Figure 6.b the set $\mathcal{G}_j$ consists of two crescents, one near the top and one near the bottom of the spectrum. In Figure 6.c the crescents are larger. In Figure 6.d we see that even after 60 iterations, the set $\mathcal{G}_j$ does not yet enclose the entire spectrum. Notice how the roots of both polynomials lie on arcs around the spectrum. The roots converge to the spectrum only as the number of iterations approaches 104.

5.5. An Elliptic Partial Differential Equation

The matrices in this example are central difference approximations to the partial differential operator

$$\mathcal{L}u = -\Delta u + \beta u_x + \gamma u$$

on the domain $[0,1] \times [0,1]$ with Dirichlet boundary conditions on a 10×10 mesh, that is, with mesh width $h = 1/11$. A conjugate gradient method was implemeted with $C = I$ and $B = A^*A$. Figure 7.a shows the spectrum and sets $\mathcal{H}(A)$ and $\mathcal{H}^{-1}(A^{-1})$ with $\beta = 30$ and $\gamma = -200$. This value of γ was chosen to keep the spectrum of A in the right half plane, but barely so. Of course, the field of values extends well into the left-hand plane. Notice that $\mathcal{H}^{-1}(A^{-1})$ is much bigger than $\mathcal{H}(A)$. In Figure 7.b, the set $\mathcal{G}_j$ is empty, while in Figure 7.c it is virtually empty. Suddenly, in the next ten iterations the set $\mathcal{G}_j$ becomes almost equal to $\mathcal{G}_n$.

Figure 8.a shows the spectrum and sets $\mathcal{H}(A)$ and $\mathcal{H}^{-1}(A^{-1})$ with $\beta = 22.5$ and $\gamma = -410$. Now, the set $\mathcal{H}^{-1}(A^{-1})$ is the exterior of the little bubble containing the origin and $\mathcal{G}_n = \mathcal{H}(A) \cap \mathcal{H}^{-1}(A^{-1})$ is two disjoint boxes. This indefinite problem poses a greater challenge to our scheme. Figures 8.b-8.c show that the roots of both polynomials lie in arcs around the spectrum for rather large values of j. Finally, in Figure 8.d the set $\mathcal{G}_j$ is beginning to grow.

6. Concluding Remarks

The basic tool used in this paper is the result that for a conjugate gradient iteration the roots of the orthogonal polynomials lie in $\mathcal{F}_B(CA)$, while the reciprocals of the roots of the residual polynomials lie in $\mathcal{F}_B((CA)^{-1})$. It was demonstrated that in some contexts the roots of these two polynomials can be used to construct regions in the complex plane that describe the spectrum of CA. A look at the examples should convince the reader that these roots provide much more information than can be derived from the sets $\mathcal{G}_j$. The sets $\mathcal{G}_j$ are a very conservative construction. Both sets of roots seem to hover around the spectrum. It seems that finer tools must be developed to make use of this information.

The conjugate gradient method, in the general case, requires a large amount of computation and storage. The work required for j iterations is proportional to j^2n and the storage required is approximately jn. For that reason, most practical implementations, such as GMRES(k) [SaSc86], are either restarted when the storage or work required becomes prohibitively large or discard past vectors and abandon optimality. From the above examples we see that, in some cases, it may be necessary to take a large number of iterations to garner much solid information from the roots. On the other hand, the lack of information may suggest that restarting may result in a divergent iteration. The relationship between the placement of the roots and the viability of a restarted method warrants further study.

Finally, we suggest that the QMR algorithm will provide orthogonal and residual polynomials that satisfy the theorems of Section 2 and will provide the

roots at a much smaller cost than the conjugate gradient method. We hope to investigate this observation in a forthcoming paper.

Acknowledgements

We would like to thank Bernd Fischer and Roland Freund for helpful discussions. In particular, we thank Roland Freund for suggesting a more complete form of the Theorems 2.1 and 2.3. We would also like to thank Steve Ashby for help with the Matlab codes that produced the figures and Martin Gutknecht for a careful reading and helpful comments. This work was funded in part by the National Science Foundation under Grant DMS–9015259 and the Department of Energy under Grant DE–FG02–90ER25086.

References

[As85] S. F. Ashby, *CHEBYCODE: A FORTRAN Implementation of Manteuffel's Adaptive Chebyshev Algorithm*, University of Illinois Dept. of Computer Science Technical Report 1203, May, 1985.

[As91] S. F. Ashby, *Minimax Polynomial Preconditioning for Hermitian Linear Systems*, SIMAX, Vol. 12, No. 4, (1991) pp. 766–789.

[AsGu91] S. F. Ashby and M. H. Gutknecht, *A Matrix Analysis of Orthogonal Error Methods*, in preparation.

[AMO91] S. F. Ashby, T. A. Manteuffel and J. S. Otto, *A Comparison of Chebychev and Least Squares Adaptive Polynomial Preconditioning for Hermitian Positive Definite Linear Systems*, SIAM J. Sci. Stat. Comput. 13 (1992), pp. 1-29.

[AMS89] S. F. Ashby, T. A. Manteuffel and P. E. Saylor, *Adaptive Polynomial Preconditioning for Hermitian Indefinite Linear Systems*, BIT, 29(4), pp. 583–609, 1989.

[AMS90] S. F. Ashby, T. A. Manteuffel and P. E. Saylor, *A Taxonomy for Conjugate Gradient Methods*, SIAM J. Numer. Anal., 26 (1990), pp. 1542-1568.

[AGG88] P. Arbenz, W. Gander and G.H. Golub, *Restricted Rank Modification of the Symmetric Eigenvalue Problem: Theoretical Considerations*, Lin. Alg. Appl., Vol 104, (1988) pp.75–95.

[ArGo88] P. Arbenz and G.H. Golub, *On the Spectral Decomposition of Hermitian Matrices Subjected to Indefinite Low Rank Perturbations*, SIAM J. Matrix Anal. Appl., Vol. 9, (1988), pp. 40–58.

[CGO76] P. Concus, G. H. Golub and D. P. O'Leary, *A Generalized Conjugate Gradient Method for the Numerical Solution of Elliptic Partial Differential Equations* in Sparse Matrix Computations, J. R. Bunch and D. J. Rose, eds., Academic Press, New York, 1976.

[Ei91] M. Eiermann, *Fields of Values and Iterative Methods*, IBM Deutschland Report, June 1991.

[FaMa84] V. Faber and T. A. Manteuffel, *Necessary and Sufficient Conditions for the Existence of a Conjugate Gradient Method*, SIAM J. Numer. Analysis, Vol. 21, No. 2, 352–362 (1984).

[FaMa87] V. Faber and T. A. Manteuffel, *Orthogonal Error Methods*, SIAM J. Numer. Analysis, Vol. 24, No. 1, 170–187 (1987).

[Fr89] R. W. Freund, *Pseudo-Ritz Values for Indefinite Hermitian Matrices*, RIACS Technical Report TR 89.33 (August 1989).

[Fr90] R. W. Freund, *On Conjugate Gradient Type Methods and Polynomial Preconditioners for a Class of Complex Non-Hermitian Linear Systems*, Numer. Math., 57 (1990), pp 285-312.

[Fr91] R. W. Freund, *Quasi-Kernel Polynomials and Their Use in Non-Hermitian Matrix Iterations*, RIACS Technical Report 91.20, October, 1991.

[FrNa91] R. W. Freund and N. M. Nachtigal, *QMR: a Quasi-Minimal Residual Method for Non-Hermitian Linear Systems*, Eidgenösslsche Technische Hochschule Zürich, IPS Research Report No. 91-05, March, 1991.

[Ga77] F.R. Gantmacher, "Matrix Theory", Chelsea Press, New York, 1977, Vol. I.

[Go73] G. H. Golub, *Some Modified Matrix Eigenvalue Problems*, SIAM Review, Vol. 15, No. 2, April, 1973, pp 318-334.

[GoVL89] G. H. Golub and C. F. Van Loan, "Matrix Computations," Johns Hopkins University Press, Baltimore, 1989.

[He62] P. Henrici, *Bounds for Iterates, Inverses, Spectral Variation and Fields of Values of Non-normal Matrices*, Numer. Math., Vol. 4, (1962) pp. 24-40.

[HeSt52] M. R. Hestenes and E. Stiefel, *Methods Of Conjugate Gradients for Solving Linear Systems*, Nat. Bureau Standards J. Res., 49 (1952), pp. 409-435.

[HoJo91] R. A. Horn and C. R. Johnson, " Topics in Matrix Analysis,", Cambridge University Press, Cambridge, 1991.

[Ho64] A. S. Householder, "The Theory of Matrices in Numerical Analysis," pp. 37–57, Blaisdell Press, New York, 1964.

[JeYo83] K. C. Jea and D. M. Young, *On the Simplification of Generalized Conjugate Gradient Methods for Nonsymmetrizable Linear Systems*, Lin. Alg. Appl., 52 (1983), pp. 399-417.

[JMP83] O. G. Johnson, C. A. Micchelli and G. Paul, *Polynomial Preconditioning for Conjugate Gradient Calculations*, SIAM J. Numer. Anal., Vol. 20, No. 2, (1983) pp. 362–376.

[Ma78] T. A. Manteuffel *Adaptive Procedures for Estimating Parameters for the Nonsymmetric Tchebychev Iteration*, Numer. Math., 31,(1978) pp. 183–208.

[NRT92] N. M. Nachtigal, L. Reichel and N. L. Trefethen, *A Hybrid GMRES Algorithm for Nonsymmetric Linear Systems*, to appear SIAM J. Matrix Analysis.

[Pa80] B. N. Parlett, "The Symmetric Eigenvalue Problem", Prentice-Hall, Englewood Cliffs, N.J., 1980.

[Sa83] Y. Saad, *Iterative Solution of Indefinite Symmetric Linear Systems by Methods Using Orthogonal Polynomials over Two Disjoint Intervals*, SIAM J. Numer. Anal., Vol. 20, No. 4, (1983) pp. 784–810.

[Sa85] Y. Saad, *Practical Use of Polynomial Preconditionings for the Conjugate Gradient Method*, SIAM J. Sci. Stat. Comput., Vol. 6, No. 4, (1985) pp. 865–881.

[SaSc86] Y. Saad and M. H. Schultz, *GMRES: a Generalized Minimal Residual Algorithm for Solving Non-Symmetric Linear Systems*, SIAM J. Sci. Stat. Comput., Vol 7, (1986), pp. 856–869.

[SmSa82] D. C. Smolarski and P. E. Saylor, *An Optimum Iterative Method for Solving Any Linear System with a Square Matrix*, BIT, Vol. 28 (1988), pp. 163–178.

[St73] G.W. Stewart, "Introduction to Matrix Computations," Academic Press, New York, 1973.

[St58] E. L. Stiefel, *Kernel Polynomials in Linear Algebra and their Numerical Applications*, U.S. Natioan Bureau of Standards, Applied Mathematics Series, Vol. 49, (1958) pp. 1–22.

[Ta68] O. Taussky, *Some Topics Concerning Bounds for Eigenvalues of Finite Matrices*, In: Survey of Numerical Analysis (J. Todd, ed.), Ch. 8, pp. 279–297, McGraw Hill, New York, 1962.

[SlVo86] A. van der Sluis and H. A. van der Vorst, *The Rate of Convergence of Conjugate Gradients*, Numer. Math., 48 (1986), pp.543-560.

[Wi65] J. H. Wilkinson, " The Algebraic Eigenvalue Problem", Clarendon Press, Oxford, 1965.

[YoJe80] D. M. Young and K. C. Jea, *Generalized Conjugate-Gradient Acceleration of Nonsymmetrizable Iterative Methods*, Lin. Alg. Appl., 34 (1980), pp. 159-194.