

IMPLEMENTATION OF THE CONJUGATE
GRADIENT METHOD USING SHORT MULTIPLE
RECURSIONS

by

Teri L. Barth

B. S., Colorado State University, 1977

M. S., University of Colorado at Denver, 1992

A thesis submitted to the
University of Colorado at Denver
in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Applied Mathematics

1996

This thesis for the Doctor of Philosophy

degree by

Teri L. Barth

has been approved

by

Thomas A. Manteuffel

William L. Briggs

Leopoldo P. Franca

Stephen F. McCormick

Thomas F. Russell

Date _____

Barth, Teri L. (Ph.D., Applied Mathematics)

Implementation Of The Conjugate Gradient
Method Using Short Multiple Recursions

Thesis directed by Professor Thomas A. Manteuffel

ABSTRACT

Iterative methods are important mechanisms for solving linear systems of equations of the form, $A\underline{x} = \underline{b}$, when A is a large sparse nonsingular matrix. When an efficient implementation exists, the conjugate gradient (CG) method is a popular technique of this type. This method is implemented via the construction of an orthogonal basis for the underlying Krylov subspace. When this basis can be constructed using recursions that involve only a few terms at each step, then a practical CG algorithm exists. The current theory from Faber and Manteuffel regarding the economical implementation of the conjugate gradient method does not take into account all possible forms of short recurrences. By considering a more general form of recursion, we will extend the class of matrices for which a practical CG algorithm is known to exist.

This study begins with a review of the conjugate gradient method, and the existing theory regarding its economical implementation. An overview of previous work concerning unitary and shifted unitary matrices will be presented since this motivates our definition of a more general form of short multiple recursion. Sufficient and partial necessary conditions on the matrix A will be determined in order that a

CG method can be implemented using this type of recursion. One sufficient condition is for the matrix A to be B -normal(ℓ, m). These are normal matrices whose adjoint can be expressed as the ratio of two polynomials in A . Another sufficient condition is for A to be a low rank perturbation of a B -normal(ℓ, m) matrix. An important example of this type is a matrix that is self-adjoint plus low rank. These results will be illustrated with a few numerical examples. In addition to extending the class of matrices for which a practical CG algorithm exists, this research opens the door to the possibility that other forms of short recurrences may exist that would admit an efficient CG algorithm for even a wider class of matrices.

This abstract accurately represents the content of the candidate's thesis. I recommend its publication.

Signed _____

Thomas A. Manteuffel

CONTENTS

Chapter

1. Introduction	1
1.1 Motivation And Organization	1
1.2 Notation	4
2. The Conjugate Gradient Method	5
2.1 Introduction	5
2.2 Derivation Of The Conjugate Gradient Method	5
2.3 Implementation	9
2.4 Economical Conjugate Gradient Algorithms	13
3. Unitary And Shifted Unitary Matrices	19
3.1 Introduction	19
3.2 A Short Double Recursion For Unitary Matrices	20
3.3 A Minimal Residual Algorithm For Shifted Unitary Matrices	22
3.4 Conjugate Gradient Algorithms For General <i>B</i> -Unitary And Shifted <i>B</i> -Unitary Matrices	25
3.5 A New Minimal Residual Algorithm For Shifted Unitary Matrices	27
4. Multiple Recursion Formulation	31
4.1 Introduction	31
4.2 <i>B</i> -Normal(ℓ, m) Matrices	33
4.3 A Multiple Recursion For <i>B</i> -Normal(ℓ, m) Matrices	39
4.4 A Multiple Recursion For Generalizations Of <i>B</i> -Normal(ℓ, m) Matrices	47
4.5 Breakdown In The Single (s, t)-Recursion	58

4.6	Concluding Remarks	60
5.	Implementation And Numerical Results	62
5.1	Introduction	62
5.2	B -Normal(ℓ, m) Matrices	62
5.3	Generalizations Of B -Normal(ℓ, m) Matrices	71
5.4	Concluding Remarks	90
6.	Necessary Conditions	91
6.1	Introduction	91
6.2	Preliminaries	92
6.3	Proof Of Necessary Conditions	112
7.	Conclusion	144
<u>Appendix</u>		
A.	Properties Of Determinants	146
<u>References</u>	148

1. Introduction

1.1 Motivation And Organization

A conjugate gradient method for the solution of an $N \times N$ linear system of equations

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

is defined with respect to an inner product matrix B , and the matrix A . It is a Krylov subspace method in which the B -norm of the error is minimized at each step (see Section 2.2). A conjugate gradient method is implemented via the construction of a B -orthogonal basis $\{\underline{p}_i\}_{i=0}^j$ for $\mathcal{K}_{j+1}(\underline{r}_0, A)$, where

$$\mathcal{K}_{j+1}(\underline{r}_0, A) = \text{sp}\{\underline{r}_0, A\underline{r}_0, \dots, A^j\underline{r}_0\},$$

is the Krylov subspace of dimension $j + 1$ generated by the initial residual \underline{r}_0 and the matrix A . The \underline{p}_i 's are called direction vectors. There are many algorithms for implementing the method, for example: Orthodir, Orthomin, Orthores, and GMRES.

If a B -orthogonal basis can be constructed with some form of a short recursion, then the work and storage requirements to implement the method are minimal, and we say an economical conjugate gradient algorithm exists. In 1984, Faber and Manteuffel [5] proved that the class of matrices for which a conjugate gradient method can be implemented via a single short $(s + 2)$ -term recursion of the form

$$\underline{p}_{j+1} = A\underline{p}_j - \sum_{i=j-s}^j \sigma_{i,j} \underline{p}_i,$$

for the direction vectors, is limited to matrices that are B -normal(s) or to matrices

with only a small number of distinct eigenvalues (see Section 2.4). It became apparent in the work done by Gragg [10] that this form of recursion is not general enough to account for all possible short recurrences. The unitary matrix is an example of a matrix for which an orthonormal basis cannot be constructed with a short $(s + 2)$ -term recursion, however, Gragg demonstrated that this basis can be constructed using a pair of recurrence formulas. Shifted unitary matrices have the form

$$A = \rho U + \xi I,$$

where U is unitary, I is the identity matrix, and ρ and ξ are complex scalars. Jagels and Reichel ([16],[17]) observed that for matrices of this form, $\mathcal{K}_{j+1}(\underline{x}_0, A) = \mathcal{K}_{j+1}(\underline{x}_0, U)$, and thus, the same double recursion as for the unitary case, can be used to construct an orthonormal basis. They continued, showing how to derive an efficient minimal residual algorithm. This motivates the work in this thesis concerning alternate forms of short recurrences. By considering a more general form of recursion, we will show that the class of matrices for which a practical conjugate gradient algorithm is known to exist, can be extended.

In Chapter 2, the conjugate gradient method is defined. We then discuss how the method is implemented. An overview of previous work concerning economical conjugate gradient algorithms is given. Chapter 3 summarizes the work done by Gragg, Jagels and Reichel on unitary and shifted unitary matrices.

The double recursion for unitary and shifted unitary matrices can be rewritten as a single (s, t) -recursion of the form

$$\underline{p}_{j+1} = A\underline{p}_j + \sum_{i=j-t}^{j-1} \beta_{i,j} A\underline{p}_i - \sum_{i=j-s}^j \sigma_{i,j} \underline{p}_i.$$

If A is unitary, $t = 1$ and $s = 0$; if A is shifted unitary, $t = 1$ and $s = 1$. We notice that if A is unitary, its adjoint, $A^* = \bar{A}^T$, can be written as the ratio of two

polynomials,

$$A^* = A^{-1} = \frac{p_0(A)}{q_1(A)},$$

where $p_0(A) = I$, and $q_1(A) = A$. The degrees of these polynomials correspond to the number of terms needed in the single (s, t) -recursion to construct an orthogonal basis.

In Chapter 4, we will define and characterize a more general class of normal matrices, called B -normal (ℓ, m) matrices. The B -adjoint of these matrices, A^\dagger , (see Section 2.4), can be expressed as the ratio of two polynomials,

$$A^\dagger = \frac{p_\ell(A)}{q_m(A)},$$

of degrees ℓ and m respectively. In the absence of a condition we call breakdown, if A is B -normal (ℓ, m) , then a B -orthogonal basis can be obtained using a single (s, t) -recursion, with $(s, t) = (\ell, m)$. This result can be extended to more general matrices, which include low rank perturbations of B -normal (ℓ, m) matrices. An important example of this type is a matrix that is self-adjoint plus low rank.

Since breakdown is possible using the single (s, t) -recursion, we will show how the problem can be reformulated as a multiple recursion that avoids the possibility of breakdown. Sufficient conditions on the matrix A are then stated in order that a B -orthogonal basis can be constructed using this form of multiple recursion.

In Chapter 5, we will discuss the details on how to implement the method using this form of recursion. Numerical examples are given comparing the results to those using a full conjugate gradient algorithm.

To determine if there are any other matrices for which a B -orthogonal basis can be constructed using this form of multiple recursion, we must answer the question of whether the sufficient conditions are also necessary. This is a much more difficult problem which we will answer in Chapter 6 for only a restricted subset of

the multiple recursions derived earlier. A brief summary is given in Chapter 7.

1.2 Notation

The notation will be explained when it is first introduced. For convenience, we summarize this below.

$\mathcal{R}^n, \mathcal{C}^n$	–	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.
$\mathcal{K}, \mathcal{V}, \dots$	–	Other calligraphic letters denote subspaces of \mathcal{R}^n or \mathcal{C}^n .
A, B, \dots	–	Upper case Roman (Greek) letters denote matrices.
$\underline{x}, \underline{p}, \dots$	–	Underlined lower case Roman (Greek) letters denote vectors.
α, β, \dots	–	Lower case Roman (Greek) letters denote scalars.
$\langle \cdot, \cdot \rangle, \ \cdot\ $	–	Euclidean inner product on \mathcal{C}^n and induced norm.
$\langle B \cdot, \cdot \rangle, \ \cdot\ _B$	–	B –inner product on \mathcal{C}^n and induced B –norm.
$-_B$	–	Represents orthogonality in the B – inner product.
A^*	–	Euclidean adjoint of A , $A^* = \bar{A}^T$.
A^\dagger	–	B –adjoint of A , $A^\dagger = B^{-1}A^*B$.
$\text{sp}\{\underline{x}_j\}$	–	The linear span of the vectors \underline{x}_j .
$\Sigma(A)$	–	Spectrum of A .
$\mathcal{K}_\ell(\underline{x}_0, A)$	–	$\text{sp}\{\underline{x}_0, A\underline{x}_0, \dots, A^{\ell-1}\underline{x}_0\}$. Krylov subspace of dimension ℓ .
$d(\underline{x}_0, A)$	–	Degree of the minimal annihilating polynomial of \underline{x}_0 .
$d(A)$	–	Degree of the minimal polynomial of A .

2. The Conjugate Gradient Method

2.1 Introduction

Consider the linear system of equations

$$A\underline{x} = \underline{b}, \tag{2.1}$$

where A is a general $N \times N$ nonsingular matrix, and \underline{x} and $\underline{b} \in \mathcal{C}^N$. This chapter is concerned with the solution of (2.1) using a conjugate gradient method.

After some background on polynomial iterative methods, a general definition of the conjugate gradient method will be given. Next, we describe how the method is implemented. Of particular interest will be the conditions under which a conjugate gradient method can be implemented with some form of a short recursion, thus requiring only minimal amounts of computation and storage. An overview of previous results on this topic is given. We conclude this chapter by noting that the current theory on economical implementation of the conjugate gradient method does not take into account all possible forms of short recursions. This motivates the further study of conjugate gradient algorithms using alternate forms of short recursions.

2.2 Derivation Of The Conjugate Gradient Method

Given an initial guess \underline{x}_0 to (2.1), an iterative method produces a sequence of approximations $\{\underline{x}_0, \underline{x}_1, \dots\}$ to the exact solution $\underline{x} = A^{-1}\underline{b}$. Iterative methods are important techniques for obtaining a solution when A is large and sparse.

Consider an iterative method where for $j = 0, 1, \dots$, the new iterate is

defined by,

$$\underline{x}_{j+1} = \underline{x}_0 + \sum_{i=0}^j \alpha_{i,j} \underline{r}_i, \quad (2.2)$$

where,

$$\underline{r}_i = \underline{b} - A\underline{x}_i = A(\underline{x} - \underline{x}_i) = A\underline{e}_i, \quad (2.3)$$

is the residual at step i , and \underline{e}_i denotes the error at step i . Each new iterate is updated with some linear combination of previous residuals. By subtracting both sides of (2.2) from the exact solution \underline{x} , and using (2.3), we obtain an expression for the error at step $j + 1$,

$$\underline{e}_{j+1} = \underline{e}_0 - \sum_{i=0}^j \alpha_{i,j} A\underline{e}_i.$$

Let I_N denote the $N \times N$ identity matrix, and note that

$$\underline{e}_0 = I_N \underline{e}_0 = p_0(A) \underline{e}_0,$$

where $p_0(0) = 1$. Suppose that for $i = 0, \dots, j$, $\underline{e}_i = p_i(A) \underline{e}_0$, and $p_i(0) = 1$. By induction, we obtain

$$\begin{aligned} \underline{e}_{j+1} &= \underline{e}_0 - \sum_{i=0}^j \alpha_{i,j} A p_i(A) \underline{e}_0 \\ &= [I_N - A \tilde{p}_j(A)] \underline{e}_0 \\ &= p_{j+1}(A) \underline{e}_0, \end{aligned} \quad (2.4)$$

for some polynomial, $p_{j+1}(\mu) = [1 - \mu \tilde{p}_j(\mu)]$, of degree $\leq j + 1$. Furthermore, we see that $p_{j+1}(0) = 1$. Notice that the residual at step $j + 1$ can be written as

$$\underline{r}_{j+1} = A p_{j+1}(A) \underline{e}_0 = p_{j+1}(A) \underline{r}_0.$$

The polynomials $p_{j+1}(\mu)$ are known as residual polynomials. Since the error at each step can be written as a polynomial in A times the initial error, iterative methods of this form are called polynomial methods.

Define the $(j + 1)$ 'st Krylov subspace generated by \underline{r}_0 and the matrix A as

$$\mathcal{K}_{j+1}(\underline{r}_0, A) = \text{sp}\{\underline{r}_0, A\underline{r}_0, \dots, A^j\underline{r}_0\}. \quad (2.5)$$

The vectors $\{A^i\underline{r}_0\}_{i=0}^j$ are called Krylov vectors.

Denote $d(A)$ as the degree of the minimal polynomial of A , that is,

$$d(A) = \min_{p: p \text{ monic}} \{\deg p : p(A) = 0\}.$$

Define $d = d(\underline{z}, A)$ to be the maximum dimension of the Krylov subspace generated by \underline{z} and the matrix A . This is the degree of the minimal annihilating polynomial of \underline{z} ,

$$d = d(\underline{z}, A) = \min_{p: p \text{ monic}} \{\deg p : p(A)\underline{z} = \underline{0}\}.$$

Since the Krylov subspaces are nested, and for every i , $\underline{r}_i = p_i(A)\underline{r}_0 \in \mathcal{K}_{i+1}(\underline{r}_0, A)$, it follows that (2.2) can be rewritten as,

$$\begin{aligned} \underline{x}_{j+1} &= \underline{x}_0 + p_j(A)\underline{r}_0 \\ &= \underline{x}_0 + \hat{\underline{z}}_j, \quad \text{where } \hat{\underline{z}}_j \in \mathcal{K}_{j+1}(\underline{r}_0, A), \end{aligned}$$

or equivalently,

$$\underline{x}_{j+1} = \underline{x}_j + \underline{z}_j,$$

for some $\underline{z}_j \in \mathcal{K}_{j+1}(\underline{r}_0, A)$. We make the following definition:

DEFINITION 2.1 A Krylov subspace method is an iterative method whose iterates are defined by:

$$\begin{aligned} \underline{x}_{j+1} &= \underline{x}_j + \underline{z}_j, & \underline{z}_j &\in \mathcal{K}_{j+1}(\underline{r}_0, A), \\ \underline{e}_{j+1} &= p_{j+1}(A)\underline{e}_0, & p_j(0) &= 1. \end{aligned}$$

There are many Krylov subspace methods. What distinguishes a method is the way in which the vector \underline{z}_j is chosen from $\mathcal{K}_{j+1}(\underline{r}_0, A)$. Since our goal is to find a solution to (2.1), a logical strategy would be to choose $\underline{z}_j \in \mathcal{K}_{j+1}(\underline{r}_0, A)$ so

that the error is as small as possible. However, we would also like to accomplish this with a relatively small amount of work. As we will see later, these two goals are not always compatible.

Since $\|\underline{e}_{j+1}\| = \|p_{j+1}(A)\underline{e}_0\| \leq \|p_{j+1}(A)\| \|\underline{e}_0\|$, we differentiate two basic strategies for making the error small:

- (1) Make $\|p_{j+1}(A)\|$ small,
- (2) make $\|p_{j+1}(A)\underline{e}_0\|$ small.

Methods that employ the first strategy involving matrix norms, are often referred to as Chebychev-like methods. They require some knowledge of the spectrum of A to implement. Methods that utilize the second strategy are called conjugate gradient-like methods. They are based upon optimization in an inner product norm, or on some orthogonalization property. No knowledge of the spectrum is required for their implementation.

As the name suggests, conjugate gradient methods utilize the second strategy. They are Krylov subspace methods in which the error is minimized at every step in an inner product norm. We will define this method with respect to a Hermitian positive definite inner product matrix B , and the system matrix A , and denote it as $\mathcal{CG}(B, A)$.

DEFINITION 2.2 A conjugate gradient method, $\mathcal{CG}(B, A)$, is a Krylov subspace method whose iterates are defined uniquely as follows:

$$\begin{aligned} \underline{x}_{j+1} &= \underline{x}_j + \underline{z}_j, & \underline{z}_j &\in \mathcal{K}_{j+1}(\underline{x}_0, A), \\ \|\underline{e}_{j+1}\|_B & & \text{is minimized over } &\mathcal{K}_{j+1}(\underline{x}_0, A). \end{aligned} \tag{2.6}$$

The second condition is equivalent to requiring

$$\underline{e}_{j+1} \perp_B \underline{z}, \quad \forall \underline{z} \in \mathcal{K}_{j+1}(\underline{x}_0, A).$$

There are many conjugate gradient methods. Consider for example, if A is symmetric positive definite, $\langle A \cdot, \cdot \rangle$ defines an inner product, and we could take

$B = A$. $\mathcal{CG}(B, A)$ minimizes the A norm of the error, which results in the original conjugate gradient method of Hestenes and Stiefel [14]. Another example of a conjugate gradient method is given by taking $B = A^*A$. Notice that since we are assuming that A is nonsingular, A^*A is Hermitian positive definite. In this case, $\mathcal{CG}(B, A)$ minimizes

$$\|\underline{e}_{j+1}\|_{A^*A} = \langle A^*A\underline{e}_{j+1}, \underline{e}_{j+1} \rangle^{\frac{1}{2}} = \langle \underline{r}_{j+1}, \underline{r}_{j+1} \rangle^{\frac{1}{2}},$$

yielding a minimal residual method. A detailed taxonomy of conjugate gradient methods can be found in [1].

2.3 Implementation

A conjugate gradient method is implemented via the construction of a B -orthogonal basis for the Krylov subspace. This follows directly from the definition of the method. To see this, first, we obtain an expression for the error at each step in terms of the error at the previous step by subtracting the first equation in (2.6) from the exact solution yielding

$$\underline{e}_{j+1} = \underline{e}_j - \underline{z}_j. \tag{2.7}$$

According to Definition 2.2, at each step $j + 1$, we must choose $\underline{z}_j \in \mathcal{K}_{j+1}(\underline{r}_0, A)$ so that

$$\underline{e}_{j+1} \perp_B \underline{z}, \quad \forall \underline{z} \in \mathcal{K}_{j+1}(\underline{r}_0, A).$$

Since $\mathcal{K}_j(\underline{r}_0, A) \subset \mathcal{K}_{j+1}(\underline{r}_0, A)$, it follows that for every $\underline{z} \in \mathcal{K}_j(\underline{r}_0, A)$,

$$\begin{aligned} 0 &= \langle B\underline{e}_{j+1}, \underline{z} \rangle \\ &= \langle B\underline{e}_j, \underline{z} \rangle - \langle B\underline{z}_j, \underline{z} \rangle, \end{aligned}$$

where the second equality results from substituting in (2.7). From the definition, it follows that for every $\underline{z} \in \mathcal{K}_j(\underline{r}_0, A)$, $\langle B\underline{e}_j, \underline{z} \rangle = 0$, which implies $\langle B\underline{z}_j, \underline{z} \rangle = 0$.

Therefore, we obtain two conditions:

$$\begin{aligned} \underline{z}_j &\in \mathcal{K}_{j+1}(\underline{x}_0, A), \quad \text{and} \\ \langle B\underline{z}_j, \underline{z} \rangle &= 0, \quad \forall \underline{z} \in \mathcal{K}_j(\underline{x}_0, A). \end{aligned}$$

These conditions say at each step, we must choose $\underline{z}_j \in \mathcal{K}_{j+1}(\underline{x}_0, A)$ such that \underline{z}_j is B -orthogonal to everything in $\mathcal{K}_j(\underline{x}_0, A)$. This means a conjugate gradient method can be implemented by constructing a B -orthogonal basis $\{\underline{p}_i\}_{i=0}^j$ for $\mathcal{K}_{j+1}(\underline{x}_0, A)$. Here, $\underline{p}_i \in \mathcal{K}_{i+1}(\underline{x}_0, A)$, and $\underline{p}_i \perp_B \mathcal{K}_i(\underline{x}_0, A)$, for $i = 0, \dots, j$. The \underline{p}_i 's are known as direction vectors. Given an initial vector, $\underline{p}_0 = \underline{x}_0$, a B -orthogonal basis is unique up to scale, or the length of the vectors. At step $j + 1$, let

$$\underline{z}_j = \alpha_j \underline{p}_j,$$

where α_j is known as the step length.

To determine the step length, notice that $\underline{p}_j \in \mathcal{K}_{j+1}(\underline{x}_0, A)$. It follows from the definition that

$$0 = \langle B\underline{e}_{j+1}, \underline{p}_j \rangle = \langle B\underline{e}_j, \underline{p}_j \rangle - \alpha_j \langle B\underline{p}_j, \underline{p}_j \rangle.$$

Since B is Hermitian positive definite, $\langle B\underline{p}_j, \underline{p}_j \rangle \neq 0$, and we can solve for α_j yielding,

$$\alpha_j = \frac{\langle B\underline{e}_j, \underline{p}_j \rangle}{\langle B\underline{p}_j, \underline{p}_j \rangle}.$$

Summarizing the above, to implement $\mathcal{CG}(B, A)$, given \underline{p}_0 , at each step, compute:

- 1) $\underline{p}_j \in \mathcal{K}_{j+1}(\underline{p}_0, A)$ such that $\underline{p}_j \perp_B \mathcal{K}_j(\underline{p}_0, A)$,
- 2) $\underline{x}_{j+1} = \underline{x}_j + \alpha_j \underline{p}_j$, $\alpha_j = \frac{\langle B\underline{e}_j, \underline{p}_j \rangle}{\langle B\underline{p}_j, \underline{p}_j \rangle}$.

Since the computation of α_j involves the unknown quantity of the error, the inner product matrix B must be chosen so that α_j is computable. For example, if A is Hermitian positive definite, we could choose $B = A$, and

$$\alpha_j = \frac{\langle \underline{x}_j, \underline{p}_j \rangle}{\langle A\underline{p}_j, \underline{p}_j \rangle}.$$

By choosing $B = A^*A$, we obtain

$$\alpha_j = \frac{\langle r_j, A\underline{p}_j \rangle}{\langle A\underline{p}_j, A\underline{p}_j \rangle}.$$

From the above, we see that the work involved in implementing a conjugate gradient method is in the construction of a B -orthogonal basis $\{\underline{p}_i\}_{i=0}^j$ for $\mathcal{K}_{j+1}(\underline{p}_0, A)$. There may be many different ways of writing the recursion for constructing such a basis. However, since this basis is unique up to scale, as long as the vectors are scaled the same, the recursions can always be rewritten in terms of a Gram-Schmidt process,

$$\begin{aligned} \underline{p}_0 &= \underline{r}_0, \\ \underline{p}_1 &= A\underline{p}_0 - \sigma_{0,0}\underline{p}_0, & \sigma_{0,0} &= \frac{\langle BA\underline{p}_0, \underline{p}_0 \rangle}{\langle B\underline{p}_0, \underline{p}_0 \rangle}, \\ &\vdots \\ \underline{p}_{j+1} &= A\underline{p}_j - \sum_{i=0}^j \sigma_{i,j}\underline{p}_i, & \sigma_{i,j} &= \frac{\langle BA\underline{p}_j, \underline{p}_i \rangle}{\langle B\underline{p}_i, \underline{p}_i \rangle}. \end{aligned} \quad (2.8)$$

Suppose that $d(\underline{p}_0, A) = N$, where N is the dimension of A . The matrix form of this recursion at step N is given by:

$$A \begin{bmatrix} \underline{p}_0 & \underline{p}_1 & \cdots & \underline{p}_{N-1} \end{bmatrix} = \begin{bmatrix} \underline{p}_0 & \underline{p}_1 & \cdots & \underline{p}_{N-1} \end{bmatrix} \begin{bmatrix} \sigma_{0,0} & \sigma_{0,1} & \cdots & \cdots & \sigma_{0,N-1} \\ & 1 & \sigma_{1,1} & & \vdots \\ & & 1 & \ddots & \vdots \\ & & & \ddots & \vdots \\ & & & & 1 & \sigma_{N-1,N-1} \end{bmatrix}, \quad (2.9)$$

which we denote as,

$$AP_N = P_N H_N.$$

In general, H_N is a full upper Hessenberg matrix. By equating the columns in the matrix equation, we obtain the individual formulas for the direction vectors given in (2.8). Note that at any given step k , after computing \underline{p}_k , the recursions can be written in matrix form as

$$AP_k = P_{k+1} H_{k+1,k}, \quad (2.10)$$

where P_k is the matrix containing the first k columns of P_N , and $H_{k+1,k}$ is the upper left $(k+1) \times k$ corner of H_N .

The various ways of writing the recursions for the direction vectors will yield different algorithms for implementing a given conjugate gradient method. We list the Orthodir algorithm below. It is an example where the direction vectors are obtained using a full Gram-Schmidt process generalized to the B -inner product.

Table 2.1: Orthodir Algorithm

$$\begin{aligned}
\underline{r}_0 &= \underline{b} - A\underline{x}_0, \\
\underline{p}_0 &= \underline{r}_0, \\
&\vdots \\
\underline{x}_{j+1} &= \underline{x}_j + \alpha_j \underline{p}_j, & \alpha_j &= \frac{\langle B\underline{e}_j, \underline{p}_j \rangle}{\langle B\underline{p}_j, \underline{p}_j \rangle}, \\
\underline{r}_{j+1} &= \underline{r}_j - \alpha_j A\underline{p}_j, \\
\underline{p}_{j+1} &= A\underline{p}_j - \sum_{i=0}^j \sigma_{i,j} \underline{p}_i, & \sigma_{i,j} &= \frac{\langle BA\underline{p}_j, \underline{p}_i \rangle}{\langle B\underline{p}_i, \underline{p}_i \rangle}
\end{aligned}$$

Notice that in the Orthodir algorithm, the recursion for the direction vector \underline{p}_{j+1} utilizes the term $A\underline{p}_j \in \mathcal{K}_{j+2}(\underline{r}_0, A)$, to bring the recursion up to the next higher dimension Krylov subspace. When BA is definite, the residuals span the Krylov subspace [1],

$$\mathcal{K}_{j+2}(\underline{r}_0, A) = \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{j+1}\} = \text{sp}\{\underline{r}_0, \underline{r}_1, \dots, \underline{r}_{j+1}\}.$$

Instead of using the term $A\underline{p}_j$ to bring us up to $\mathcal{K}_{j+2}(\underline{r}_0, A)$, we can use \underline{r}_{j+1} . An example of an algorithm that uses this approach is the Orthomin algorithm, given in Table 2.2.

Both of these algorithms utilize all the previous direction vectors to construct \underline{p}_{j+1} . In the next section, we will see that for certain kinds of matrices A , a B -orthogonal basis can be constructed with a short recursion that uses only a few previous direction vectors.

Table 2.2: Orthomin Algorithm

$$\begin{aligned}
 \underline{r}_0 &= \underline{b} - A\underline{x}_0, \\
 \underline{p}_0 &= \underline{r}_0, \\
 &\vdots \\
 \underline{x}_{j+1} &= \underline{x}_j + \alpha_j \underline{p}_j, & \alpha_j &= \frac{\langle B\underline{e}_j, \underline{p}_j \rangle}{\langle B\underline{p}_j, \underline{p}_j \rangle}, \\
 \underline{r}_{j+1} &= \underline{r}_j - \alpha_j A\underline{p}_j, \\
 \underline{p}_{j+1} &= \underline{r}_{j+1} - \sum_{i=0}^j \sigma_{i,j} \underline{p}_i, & \sigma_{i,j} &= \frac{\langle B\underline{r}_{j+1}, \underline{p}_i \rangle}{\langle B\underline{p}_i, \underline{p}_i \rangle}
 \end{aligned}$$

2.4 Economical Conjugate Gradient Algorithms

Before discussing when a conjugate gradient method can be implemented with a short recursion for the direction vectors, we review some terminology on normal matrices.

First, the concept of an adjoint of a matrix A is generalized to the B -inner product. The B -adjoint of A is the unique matrix A^\dagger satisfying,

$$\langle BA\underline{x}, \underline{y} \rangle = \langle B\underline{x}, A^\dagger \underline{y} \rangle, \quad \forall \underline{x}, \underline{y} \in C^N.$$

This yields, $A^\dagger = B^{-1}A^*B$.

The following conditions are generalizations of those for normal matrices.

A matrix A is B -normal if it satisfies the following equivalent conditions:

- (1) $AA^\dagger = A^\dagger A$,
- (2) A and A^\dagger have the same complete set of B -orthogonal eigenvectors,
- (3) A^\dagger can be written as a polynomial of some degree in the matrix A .

We say that A is B -normal(s), if it is B -normal, and s is the degree of the polynomial, $p_s(A)$, of smallest degree for which $A^\dagger = p_s(A)$. For example, if A is B -normal(1), then A^\dagger can be written as a first degree polynomial in A .

Recall the matrix equation in (2.9) for the construction of a B -orthogonal

basis. The entries of the Hessenberg matrix H_N are given by,

$$h_{i,j} = \begin{cases} \sigma_{i,j} = \frac{\langle BA\underline{p}_j, \underline{p}_i \rangle}{\langle B\underline{p}_i, \underline{p}_i \rangle} = \frac{\langle B\underline{p}_j, A^\dagger \underline{p}_i \rangle}{\langle B\underline{p}_i, \underline{p}_i \rangle}, & \text{if } j \geq i \\ 1 & \text{if } j = i - 1 \\ 0 & \text{if } j < i - 1 \end{cases}$$

If A is B -normal(1), $A^\dagger = p_1(A)$, and thus

$$A^\dagger \underline{p}_i \in \mathcal{K}_{i+2}(\underline{p}_0, A) = \text{sp}\{\underline{p}_0, \dots, \underline{p}_{i+1}\}.$$

It follows that if $j > i + 1$,

$$\underline{p}_j \perp_B \mathcal{K}_{i+2}(\underline{p}_0, A),$$

which means $\sigma_{i,j} = 0$, yielding a tridiagonal matrix H_N . The recursion in (2.8) naturally truncates to the 3-term recursion,

$$\underline{p}_{j+1} = A\underline{p}_j - \sum_{i=j-1}^j \sigma_{i,j} \underline{p}_i.$$

A special case of B -normal(1) matrices are B -self-adjoint matrices. They satisfy:

$$\begin{aligned} A^\dagger &= A, \text{ or} \\ BA &= (BA)^* = A^*B. \end{aligned}$$

Consider an $(s + 2)$ -term recursion of the form:

$$\underline{p}_{j+1} = A\underline{p}_j - \sum_{i=j-s}^j \sigma_{i,j} \underline{p}_i, \quad \sigma_{i,j} = \frac{\langle BA\underline{p}_j, \underline{p}_i \rangle}{\langle B\underline{p}_i, \underline{p}_i \rangle}. \quad (2.11)$$

DEFINITION 2.3 The matrix A is in the class $\mathcal{CG}(s)$ if for every \underline{r}_0 , a single $(s + 2)$ -term recursion of the form given in (2.11) can be used to construct a B -orthogonal basis for $\mathcal{K}_d(\underline{r}_0, A)$.

THEOREM 2.1 $A \in \mathcal{CG}(s)$ if and only if

$$d(A) \leq s + 2 \text{ or } A \text{ is } B\text{-normal}(s).$$

Proof: The proof is a result from Faber and Manteuffel [5]. \square

This result was extended in 1988 by Joubert and Young [18], who showed that the recursions for the direction vectors in the Orthomin implementation of $\mathcal{CG}(B, A)$ also naturally truncate for B -normal(s) matrices.

In their 1984 paper, Faber and Manteuffel characterized B -normal(s) matrices. These results are given in the following lemma.

LEMMA 2.2 If A is B -normal(s);

- 1). if $s > 1$, then $d(A) \leq s^2$,
- 2). if $s = 1$, then either $d(A) = 1$, $A = A^\dagger$, or

$$A = e^{i\theta} \left(i\frac{r}{2}I + G \right),$$

where $i = \sqrt{-1}$, $r \geq 0$ is a real number, $0 \leq \theta \leq 2\pi$, and $G = G^\dagger$.

Proof: The proof is given in [5]. \square

Summarizing, if A is B -normal(s), and $s > 1$, then A has at most s^2 distinct eigenvalues. If $s = 1$ and A has more than 1 distinct eigenvalue, then all the eigenvalues of A lie on some straight line in the complex plane. Note that in the case that A is B -self adjoint, all the eigenvalues of A are real. Otherwise, the eigenvalues of A can be obtained by shifting and then rotating the eigenvalues of a B -self adjoint matrix G , yielding collinear eigenvalues.

These results say that an economical recursion for the direction vectors, of the form given by (2.11), can only be applied to a very small class of matrices. In Tables 2.3 and 2.4 we list two algorithms that are applicable when A is B -normal(1). The Omin algorithm carries the further restriction that BA be definite.

We might ask if there are other forms of short recursions that can't be

Table 2.3: Odir Algorithm

$$\begin{aligned}
\underline{r}_0 &= \underline{b} - A\underline{x}_0, \\
\underline{p}_0 &= \underline{r}_0, \\
&\vdots \\
\underline{x}_{j+1} &= \underline{x}_j + \alpha_j \underline{p}_j, & \alpha_j &= \frac{\langle B\underline{e}_j, \underline{p}_j \rangle}{\langle B\underline{p}_j, \underline{p}_j \rangle}, \\
\underline{r}_{j+1} &= \underline{r}_j - \alpha_j A\underline{p}_j, \\
\underline{p}_{j+1} &= A\underline{p}_j - \sigma_{j-1,j} \underline{p}_{j-1} - \sigma_{j,j} \underline{p}_j, & \sigma_{j-1,j} &= \frac{\langle BA\underline{p}_j, \underline{p}_{j-1} \rangle}{\langle B\underline{p}_{j-1}, \underline{p}_{j-1} \rangle}, & \sigma_{j,j} &= \frac{\langle BA\underline{p}_j, \underline{p}_j \rangle}{\langle B\underline{p}_j, \underline{p}_j \rangle}
\end{aligned}$$

written as a short $(s + 2)$ -term recursion of the form given in (2.11). Instead of computing one recursion at each step for the new direction vector, consider the multiple recursion:

$$\begin{aligned}
\underline{p}_{j+1_1} &= \sum_{i=1}^t \alpha_{i,j}^{(1)} A\underline{p}_{j_i} - \sum_{i=1}^t \beta_{i,j}^{(1)} \underline{p}_{j_i}, \\
\underline{p}_{j+1_2} &= \sum_{i=1}^t \alpha_{i,j}^{(2)} A\underline{p}_{j_i} - \sum_{i=1}^t \beta_{i,j}^{(2)} \underline{p}_{j_i}, \\
&\vdots \\
\underline{p}_{j+1_t} &= \sum_{i=1}^t \alpha_{i,j}^{(t)} A\underline{p}_{j_i} - \sum_{i=1}^t \beta_{i,j}^{(t)} \underline{p}_{j_i},
\end{aligned} \tag{2.12}$$

where \underline{p}_{j+1_i} denotes the new direction vector at this step, and \underline{p}_{j+1_i} , for $i = 2, \dots, t$ denotes $t - 1$ auxiliary vectors. Each recursion involves the vectors $\{\underline{p}_{j_i}\}_{i=1}^t$ from the previous step, and their products with A , $\{A\underline{p}_{j_i}\}_{i=1}^t$. By denoting the matrix

$$Q_j = \left[\underline{p}_{j_1}, \underline{p}_{j_2}, \dots, \underline{p}_{j_t} \right],$$

the recursions at each step $j + 1$ can be written as,

$$Q_{j+1} = AQ_jR - Q_jS, \tag{2.13}$$

where R and S are $t \times t$ matrices containing the recursion coefficients. Faber and Manteuffel conjectured that any such recursion can be rewritten in the form (2.11), for some $s \geq t$. However, we will show below that there are short multiple recursions

Table 2.4: Omin Algorithm

$$\begin{aligned}
 \underline{r}_0 &= \underline{b} - A\underline{x}_0, \\
 \underline{p}_0 &= \underline{r}_0, \\
 &\vdots \\
 \underline{x}_{j+1} &= \underline{x}_j + \alpha_j \underline{p}_j, & \alpha_j &= \frac{\langle B\underline{r}_j, \underline{p}_j \rangle}{\langle B\underline{p}_j, \underline{p}_j \rangle}, \\
 \underline{r}_{j+1} &= \underline{r}_j - \alpha_j A\underline{p}_j, \\
 \underline{p}_{j+1} &= \underline{r}_{j+1} - \beta_j \underline{p}_j, & \beta_j &= \frac{\langle B\underline{r}_{j+1}, \underline{p}_j \rangle}{\langle B\underline{p}_j, \underline{p}_j \rangle}
 \end{aligned}$$

of the form (2.13) that can't be rewritten as a single short $(s + 2)$ -term recursion of the form (2.11). More importantly, we will show that the class of matrices for which these recursions admit a conjugate gradient method, is wider than the class described by Faber and Manteuffel [5].

For any matrix A , the direction vectors can be computed using a full Gram-Schmidt process. After N steps, this may result in a dense upper Hessenberg matrix H_N . We note here, that given \underline{p}_0 , that H_N is unique. If H_N can be factored as

$$H_N = T_N B_N,$$

where B_N is nonsingular, B_N^{-1} is banded upper triangular, and T_N is banded upper Hessenberg, then the matrix equations (2.9) can be rewritten in the form,

$$AP_N B_N^{-1} = P_N T_N,$$

or,

$$AP_N \begin{bmatrix} * & \cdots & * & & \\ & \ddots & & \ddots & \\ & & \ddots & & * \\ & & & \ddots & \vdots \\ & & & & * \end{bmatrix} = P_N \begin{bmatrix} * & \cdots & * & & \\ * & \ddots & & \ddots & \\ & \ddots & \ddots & & * \\ & & \ddots & \ddots & \vdots \\ & & & * & * \end{bmatrix},$$

where the $*$'s denote possible nonzero entries. By equating the columns in the above

matrix equation, we obtain an equation at each step, of the form,

$$\delta_{j+1} \underline{p}_{j+1} = \sum_{i=j-t}^j \beta_{i,j} A \underline{p}_i - \sum_{i=j-s}^j \sigma_{i,j} \underline{p}_i. \quad (2.14)$$

In the next chapter, we will see that the unitary matrix is an example of a matrix whose direction vectors can be computed using either a short multiple recursion of the form (2.12), or a short single recursion of the form (2.14). The multiple recursion for unitary matrices is a result from Gragg, who showed the process for constructing an orthonormal basis simplifies for isometric operators, yielding a short double recursion for the direction vectors. Jagels and Reichel extended these results. By using Gragg's algorithm, they showed how to construct an efficient minimal residual algorithm for unitary and shifted unitary matrices (see Section 3.3). Although the unitary matrix is normal, it is not normal(s) for some small degree s , and thus, no short recursion of the form (2.11) exists. By considering other recursions, we will to extend the class of matrices for which we know a practical conjugate gradient algorithm exists.

3. Unitary And Shifted Unitary Matrices

3.1 Introduction

In the previous chapter, we saw that a practical implementation of a conjugate gradient method depended on the construction of a B -orthogonal basis for the Krylov subspace. If this basis can be computed with a short recursion of some form, then an economical conjugate gradient algorithm exists.

In 1984, Faber and Manteuffel [5] determined the class of matrices for which a B -orthogonal basis $\{\underline{p}_j\}_{j=0}^{d-1}$ for $\mathcal{K}_d(\underline{x}_0, A)$ could be constructed with a single $(s+2)$ -term recursion of the form

$$\underline{p}_{j+1} = A\underline{p}_j - \sum_{i=j-s}^j \sigma_{i,j} \underline{p}_i.$$

This class of matrices is limited to matrices that are B -normal(s), or to matrices with only a small number of distinct eigenvalues.

That the above recursion is not general enough to account for all possible short recursions, became apparent in the work done by Gragg [10] on unitary matrices. Gragg showed that if U is unitary, then an orthonormal basis for the Krylov subspace can be constructed using two short recursions at each step. Jagels and Reichel ([16],[17]) extended this work to obtain an economical minimal residual method for unitary and shifted unitary matrices.

A shifted unitary matrix has the form

$$\rho U + \xi I, \tag{3.1}$$

where U is a unitary matrix, I is the identity matrix, and ρ and ξ are complex scalars. It is easy to verify that these matrices are normal by noting that $AA^* = A^*A$. If

$\hat{\lambda}_j$ is an eigenvalue of U , then $\rho\hat{\lambda}_j + \xi$ is an eigenvalue of A . This means that the eigenvalues of shifted unitary matrices lie on a circle in the complex plane. Although unitary and shifted unitary matrices are normal matrices, they are not normal(s) for some small degree s , and thus, a short $(s + 2)$ -term recursion of the form (2.11) cannot be used to construct an orthonormal basis.

The results from Gragg, Jagels and Reichel, will be extended to matrices that are unitary with respect to any inner product, $\langle B\cdot, \cdot \rangle$. We conclude this chapter by presenting another algorithm for a minimal residual method for shifted unitary matrices.

3.2 A Short Double Recursion For Unitary Matrices

Suppose that U is an $N \times N$ unitary matrix, and $d = d(\underline{x}_0, U) = N$. It was shown by Gragg [10] that an orthonormal basis for $\mathcal{K}_N(\underline{x}_0, U)$ can be constructed with two short recursions. To see this, we first consider a Gram-Schmidt process to construct an orthonormal basis $\{\underline{p}_j\}_{j=0}^{N-1}$ for $\mathcal{K}_N(\underline{x}_0, U)$. Note that this is the same process as given in (2.8) with $B = I$, except the vectors are normalized to have length 1. The recursions for the \underline{p}_j 's may be written in matrix notation as

$$UP_N = P_N H_N, \tag{3.2}$$

where H_N is the $N \times N$ upper Hessenberg matrix.

Since U is unitary, and by construction P_N is unitary, it follows from (3.2) that $P_N H_N$ is unitary and

$$(P_N H_N)^* P_N H_N = H_N^* H_N = I_N,$$

where I_N is the $N \times N$ identity matrix. Therefore, H_N is a unitary, upper Hessenberg matrix. A QR factorization of H_N yields

$$H_N = Q_N R_N,$$

where Q_N is $N \times N$ unitary, and R_N is $N \times N$ upper triangular. Since $R_N = Q_N^* H_N$, it follows that R_N is also unitary. R_N being unitary and upper triangular implies that

$$R_N = \text{diag}(\dots e^{i\theta_j} \dots),$$

for some θ_j , $j = 1, \dots, N$.

The upper Hessenberg structure of H_N allows for the QR factorization to be computed efficiently using elementary unitary, or Givens matrices,

$$H_N = G_1 G_2 \cdots G_{N-1} R_N, \quad (3.3)$$

where,

$$G_j = \begin{bmatrix} I_{j-1} & & & & \\ & -\gamma_j & \sigma_j & & \\ & & \sigma_j & \bar{\gamma}_j & \\ & & & & I_{N-j-1} \end{bmatrix},$$

and $\gamma_j \in \mathcal{C}$, $|\gamma_j| \leq 1$, and $\sigma_j = (1 - |\gamma_j|^2)^{1/2}$. Since each G_j is a rank 2 correction of the identity matrix, it follows that the information content of H_N is of order N . By substituting (3.3) into (3.2) and equating columns, we obtain:

$$\begin{aligned} U_{\underline{p}_0} &= -\gamma_1 \underline{p}_0 + \sigma_1 \underline{p}_1, \\ \gamma_1 &= -\langle U_{\underline{p}_0}, \underline{p}_0 \rangle, \\ \sigma_1 &= (1 - |\gamma_1|^2)^{1/2}, \\ U_{\underline{p}_1} &= -\gamma_2 (\sigma_1 \underline{p}_0 + \bar{\gamma}_1 \underline{p}_1) + \sigma_2 \underline{p}_2, \\ \gamma_2 &= -\langle U_{\underline{p}_1}, \sigma_1 \underline{p}_0 + \bar{\gamma}_1 \underline{p}_1 \rangle, \\ \sigma_2 &= (1 - |\gamma_2|^2)^{1/2}, \\ &\vdots \end{aligned}$$

Rearranging yields the Gragg algorithm for constructing an orthonormal basis for unitary matrices. This algorithm is given in Table 3.1.

See ([10],[16],[17]) for a more complete presentation. Another derivation of this double recursion for unitary matrices is given by Watkins [24].

Table 3.1: The Gragg Algorithm for unitary matrices

$$\begin{aligned}
\underline{p}_0 &= \tilde{\underline{p}}_0 = \frac{\underline{x}_0}{\|\underline{x}_0\|}, \\
\gamma_1 &= -\langle U\underline{p}_0, \tilde{\underline{p}}_0 \rangle, \\
\sigma_1 &= (1 - |\gamma_1|^2)^{\frac{1}{2}}, \\
\sigma_1 \underline{p}_1 &= U\underline{p}_0 + \gamma_1 \tilde{\underline{p}}_0, \\
\tilde{\underline{p}}_1 &= \sigma_1 \tilde{\underline{p}}_0 + \tilde{\gamma}_1 \underline{p}_1, \\
&\vdots \\
\gamma_{j+1} &= -\langle U\underline{p}_j, \tilde{\underline{p}}_j \rangle, \\
\sigma_{j+1} &= (1 - |\gamma_{j+1}|^2)^{1/2}, \\
\sigma_{j+1} \underline{p}_{j+1} &= U\underline{p}_j + \gamma_{j+1} \tilde{\underline{p}}_j, \\
\tilde{\underline{p}}_{j+1} &= \sigma_{j+1} \tilde{\underline{p}}_j + \tilde{\gamma}_{j+1} \underline{p}_{j+1}.
\end{aligned}$$

3.3 A Minimal Residual Algorithm For Shifted Unitary Matrices

Jagels and Reichel ([16],[17]) use the Gragg algorithm to construct an efficient minimal residual algorithm for the solution of the linear system, $A\underline{x} = \underline{b}$, where A is a shifted unitary matrix (3.1). They observe that $\mathcal{K}_N(\underline{x}_0, A) = \mathcal{K}_N(\underline{x}_0, U)$. Thus, the Gragg algorithm can be used to construct an orthonormal basis for $\mathcal{K}_N(\underline{x}_0, A)$. This basis is then used to construct a minimal residual algorithm as follows:

At step k in the construction of the orthonormal basis using the Gragg algorithm, the recursions for the basis vectors, $\{\underline{p}_i\}_{i=0}^k$, can be written in matrix form as

$$UP_k = P_{k+1}H_{k+1,k}^{(U)}, \quad (3.4)$$

where $H_{k+1,k}^{(U)}$ is the $(k+1) \times k$ upper Hessenberg matrix,

$$H_{k+1,k}^{(U)} = G_1 G_2 \cdots G_{k-1} \tilde{G}_k,$$

where for $i = 1, \dots, k-1$,

$$G_i = \begin{bmatrix} I_{i-1} & & & & \\ & -\gamma_i & \sigma_i & & \\ & & \sigma_i & \tilde{\gamma}_i & \\ & & & & I_{k-i-1} \end{bmatrix}, \quad \text{and} \quad \tilde{G}_k = \begin{bmatrix} I_{k-1} & & \\ & \gamma_k & \\ & & \sigma_k \end{bmatrix}.$$

Using (3.1) and (3.4) we see that

$$AP_k = (\rho U + \xi I)P_k = P_{k+1}(\rho H_{k+1,k}^{(U)} + \xi \hat{I}_k) = P_{k+1}H_{k+1,k}^{(A)},$$

where,

$$\hat{I}_k = \begin{bmatrix} I_k \\ \underline{0}^T \end{bmatrix} \in \mathcal{C}^{(k+1) \times k}.$$

We remark here, that instead of deriving $H_{k+1,k}^{(A)}$ from $H_{k+1,k}^{(U)}$, one could obtain $H_{k+1,k}^{(A)}$ directly by substituting $U = \frac{1}{\rho}(A - \xi I)$ into the Gragg algorithm. Now, the standard argument for constructing a GMRES algorithm (c.f.[23]) is employed.

Since $\|\underline{r}_0\|_{\underline{p}_0} = r_0$, we have for some $\underline{y}_k \in \mathcal{C}^k$

$$\underline{r}_k = \underline{r}_0 - AP_k \underline{y}_k = P_{k+1} \left[\|\underline{r}_0\|_{\underline{\epsilon}_1} - H_{k+1,k}^{(A)} \underline{y}_k \right],$$

where $\underline{\epsilon}_1 = [1, 0, \dots, 0]^T$. The objective is to choose \underline{y}_k to minimize $\|\underline{r}_k\|$. Since P_{k+1} has orthonormal columns, this may be accomplished by solving the $(k+1) \times k$ least squares problem

$$\|\underline{r}_0\|_{\underline{\epsilon}_1} \approx H_{k+1,k}^{(A)} \underline{y}_k.$$

Since $(\rho H_{k+1,k}^{(U)} + \xi \hat{I}_k) = H_{k+1,k}^{(A)}$ is upper Hessenberg, a QR factorization of $H_{k+1,k}^{(A)}$ is accomplished by Givens matrices. The relationship between $H_{k+1,k}^{(A)}$ and $H_{k+1,k}^{(U)}$ allows the least squares problem to be solved using an algorithm that involves five short recursions. The algorithm is given in Table 3.2. We note here, that the \underline{v}_j 's in this algorithm are the same as our \underline{p}_j 's. See ([16],[17]) for details.

Table 3.2: Minimal Residual Algorithm: $A = \rho U + \zeta I$ (Jagels and Reichel)

Input: \underline{x}_0 , $\underline{r}_0 = \underline{b} - A\underline{x}_0$, ϵ ;

$\delta_0 = \|\underline{r}_0\|$; $\hat{\varphi}_1 = 1/\delta_0$; $\hat{\tau}_1 = \delta_0/\rho$; $\underline{w}_{-1} = \underline{p}_{-1} = \underline{v}_0 = \underline{0}$;
 $\varphi_0 = s_0 = \lambda_0 = r_{0,-1} = 0$; $r_{0,0} = \gamma_0 = \sigma_0 = c_0 = 1$;
 $\underline{v}_1 = \tilde{\underline{v}}_1 = \underline{r}_0/\delta_0$;

for $m = 1, 2, \dots$ **until** $|\hat{\tau}_{m+1}| \leq \epsilon$,

$\underline{u} = U\underline{v}_m$;

$\gamma_m = -\tilde{\underline{v}}_m^* \underline{u}$; $\sigma_m = ((1 - |\gamma_m|)(1 + |\gamma_m|))^{1/2}$;

$\alpha_m = -\gamma_m \delta_{m-1}$;

$r_{m-1,m} = \alpha_m \varphi_{m-1} + s_{m-1} \zeta/\rho$; $\hat{r}_{m,m} = \alpha_m \hat{\varphi}_m + \bar{c}_{m-1} \zeta/\rho$;

$\bar{c}_m = \hat{r}_{m,m}/(|\hat{r}_{m,m}|^2 + |\sigma_m|^2)^{1/2}$; $s_m = -\sigma_m/(|\hat{r}_{m,m}|^2 + |\sigma_m|^2)^{1/2}$;

$r_{m,m} = -c_m \hat{r}_{m,m} + s_m \sigma_m$;

$\tau_m = -c_m \hat{\tau}_m$; $\hat{\tau}_{m+1} = s_m \hat{\tau}_m$;

$\eta_m = \tau_m/r_{m,m}$; $\kappa_{m-1} = r_{m-1,m}/r_{m-1,m-1}$;

$\underline{w}_{m-1} = \alpha_m \underline{p}_{m-2} - (\underline{w}_{m-2} - \underline{v}_{m-1})\kappa_{m-1}$;

$\underline{p}_{m-1} = \underline{p}_{m-2} - (\underline{w}_{m-2} - \underline{v}_{m-1})\lambda_{m-1}$;

$\underline{x}_m = \underline{x}_{m-1} - (\underline{w}_{m-1} - \underline{v}_m)\eta_m$;

if $\sigma_m = 0$ **then** $\underline{x} = \underline{x}_m$; **exit**.

$\delta_m = \delta_{m-1}\sigma_m$; $\varphi_m = -c_m \hat{\varphi}_m + s_m \bar{\gamma}_m/\delta_m$; $\lambda_m = \varphi_m/r_{m,m}$;

$\hat{\varphi}_{m+1} = s_m \hat{\varphi}_m + \bar{c}_m \bar{\gamma}_m/\delta_m$;

$\underline{v}_{m+1} = \sigma_m^{-1}(\underline{u} + \gamma_m \tilde{\underline{v}}_m)$;

$\tilde{\underline{v}}_{m+1} = \sigma_m \tilde{\underline{v}}_m + \bar{\gamma}_m \underline{v}_{m+1}$;

$\underline{x} = \underline{x}_m$.

3.4 Conjugate Gradient Algorithms For General B -Unitary And Shifted Shifted B -Unitary Matrices

A matrix U_B is B -unitary, or isometric with respect to the inner product, $\langle B\cdot, \cdot \rangle$, if

$$\langle BU_B\underline{x}, U_B\underline{y} \rangle = \langle B\underline{x}, \underline{y} \rangle, \quad \forall \underline{x}, \underline{y} \in \mathcal{C}^N. \quad (3.5)$$

Gragg [10] gave a more general result than what we presented in Section 3.2. This result says that if a matrix U_B is isometric with respect to any inner product, $\langle B\cdot, \cdot \rangle$, then a B -orthonormal basis for $\mathcal{K}_N(\underline{r}_0, U_B)$ can be constructed with a short double recursion of the form given in Table 3.1. The derivation of this recursion is similar to that for the unitary case.

After step N of a Gram-Schmidt process which constructs a B -orthonormal basis $\{\underline{p}_i\}_{i=0}^{N-1}$ for $\mathcal{K}_N(\underline{r}_0, U_B)$ we have

$$U_B P_N = P_N H_N, \quad (3.6)$$

where H_N is an $N \times N$ upper Hessenberg matrix. Multiplying through on the left by $P_N^* U_B^* B$ yields,

$$P_N^* U_B^* B U_B P_N = P_N^* U_B^* B P_N H_N. \quad (3.7)$$

Since U_B is B -unitary, $U_B^* B U_B = B$. Using (3.6) it follows that $P_N^* U_B^* = H_N^* P_N^*$. Substituting these quantities into (3.7) yields

$$P_N^* B P_N = H_N^* P_N^* B P_N H_N.$$

Since $P_N^* B P_N = I_N$ by construction, it follows that H_N is I -unitary. Therefore, H_N can be written as a product of elementary unitary matrices. Analogous to the unitary case in Section 3.2, the substitution of this factorization for H_N into (3.6), and equating the columns, yields the recursion formulas for the \underline{p}_i 's. The iteration to produce a B -orthonormal basis is exactly like the Gragg algorithm except the standard inner product $\langle \cdot, \cdot \rangle$ is replaced by $\langle B\cdot, \cdot \rangle$. The algorithm is given in Table 3.3.

Table 3.3: The Gragg algorithm for isometric operators

$$\begin{aligned}
\underline{p}_0 &= \tilde{\underline{p}}_0 = \underline{x}_0 / \|\underline{x}_0\|_B, \\
\gamma_1 &= -\langle BU_B \underline{p}_0, \tilde{\underline{p}}_0 \rangle, \\
\sigma_1 &= (1 - |\gamma_1|^2)^{\frac{1}{2}}, \\
\sigma_1 \underline{p}_1 &= U_B \underline{p}_0 + \gamma_1 \tilde{\underline{p}}_0, \\
\tilde{\underline{p}}_1 &= \sigma_1 \tilde{\underline{p}}_0 + \bar{\gamma}_1 \underline{p}_1, \\
&\vdots \\
\gamma_{j+1} &= -\langle BU_B \underline{p}_j, \tilde{\underline{p}}_j \rangle, \\
\sigma_{j+1} &= (1 - |\gamma_{j+1}|^2)^{\frac{1}{2}}, \\
\sigma_{j+1} \underline{p}_{j+1} &= U_B \underline{p}_j + \gamma_{j+1} \tilde{\underline{p}}_j, \\
\tilde{\underline{p}}_{j+1} &= \sigma_{j+1} \tilde{\underline{p}}_j + \bar{\gamma}_{j+1} \underline{p}_{j+1}.
\end{aligned}$$

From Chapter 2, we know that an algorithm for $\mathcal{CG}(B, U_B)$, can be obtained as follows: Construct a B -orthonormal basis for $\mathcal{K}_N(\underline{x}_0, U_B)$ using the the algorithm given in Table 3.3. At each step, perform the additional recursions,

$$\begin{aligned}
\underline{x}_{j+1} &= \underline{x}_j + \alpha_j \underline{p}_j, \quad \alpha_j = \frac{\langle B \underline{e}_j, \underline{p}_j \rangle}{\langle B \underline{p}_j, \underline{p}_j \rangle} = \langle B \underline{e}_j, \underline{p}_j \rangle, \\
\underline{r}_{j+1} &= \underline{r}_j - \alpha_j A \underline{p}_j.
\end{aligned}$$

Notice that since the computation of the α_j 's involves the unknown quantity of the error, B must be chosen so that α_j is computable.

A shifted B -unitary matrix has the form

$$A = \rho U_B + \xi I.$$

Since $\mathcal{K}_N(\underline{x}_0, A) = \mathcal{K}_N(\underline{x}_0, U_B)$, the same B -orthonormal basis for $\mathcal{K}_N(\underline{x}_0, U_B)$ can

be used for $\mathcal{K}_N(\underline{x}_0, A)$. Thus a conjugate gradient method for shifted B -unitary matrices can also be implemented with short recursions.

3.5 A New Minimal Residual Algorithm For Shifted Unitary Matrices

In this section, we present an alternative minimal residual algorithm for shifted unitary matrices. This reformulation will motivate the development in Chapter 4. First, let us restate the minimal residual method for the solution of the linear system, $A\underline{x} = \underline{b}$, as a conjugate gradient method $\mathcal{CG}(B, A)$, with $B = A^*A$. The iterates produced by this method are uniquely defined by:

$$\begin{aligned} \underline{x}_{j+1} &= \underline{x}_j + \alpha_j \underline{p}_j, & \underline{p}_j &\in \mathcal{K}_{j+1}(\underline{x}_0, A), \\ \underline{e}_{j+1} &= -A^*A \underline{e}_j + \beta_j \underline{p}_j, & & \end{aligned}$$

In this context, it is clear that the direction vectors, \underline{p}_j 's, must satisfy

$$\begin{aligned} \underline{p}_j &\in \mathcal{K}_{j+1}(\underline{x}_0, A), \\ \underline{p}_j &\perp_{-A^*A} \mathcal{K}_j(\underline{x}_0, A). \end{aligned}$$

Suppose that $d(\underline{x}_0, A) = N$. We seek a basis, $\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{N-1}\}$, for $\mathcal{K}_N(\underline{x}_0, A)$, such that

$$\langle A^*A \underline{p}_j, \underline{p}_i \rangle = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases}$$

Now, suppose that A is shifted unitary. Note that in this case $\mathcal{K}_{j+1}(\underline{x}_0, A) = \mathcal{K}_{j+1}(\underline{x}_0, U)$. Thus, it is sufficient to find an A^*A orthogonal basis for $\mathcal{K}_{j+1}(\underline{x}_0, U)$. Next, note that if U is unitary with respect to I , and $A = \rho U + \xi I$, then

$$A^*AU = UA^*A. \tag{3.8}$$

It follows that U is A^*A -unitary. In other words, U is isometric with respect to the inner product $\langle A^*A \cdot, \cdot \rangle$, that is,

$$\langle A^*AU\underline{x}, U\underline{y} \rangle = \langle UA^*A\underline{x}, U\underline{y} \rangle = \langle A^*A\underline{x}, \underline{y} \rangle, \quad \forall \underline{x}, \underline{y} \in C^N.$$

From Section 3.4 it follows that the algorithm given in Table 3.3, with $B = A^*A$ and $U_B = U$, can be used to construct an A^*A -orthonormal basis. Given \underline{x}_0 , the basis thus computed can be used to solve the linear system, $A\underline{x} = \underline{b}$, by adding at each step the recursions,

$$\begin{aligned}\underline{x}_{j+1} &= \underline{x}_j + \alpha_j \underline{p}_j, & \alpha_j &= \frac{\langle A^* A \underline{e}_j, \underline{p}_j \rangle}{\langle A^* A \underline{p}_j, \underline{p}_j \rangle} = \langle \underline{r}_j, A \underline{p}_j \rangle, \\ \underline{r}_{j+1} &= \underline{r}_j - \alpha_j A \underline{p}_j.\end{aligned}$$

Note that the algorithm above requires multiplications by both U and A . To rearrange the algorithm to one requiring only multiplication by U , define the quantities

$$\underline{q}_j = A \underline{p}_j \quad \text{and} \quad \tilde{\underline{q}}_j = A \tilde{\underline{p}}_j.$$

Notice that for $A = \rho U + \xi I$, A and U commute. By multiplying the equations for \underline{p}_j , and $\tilde{\underline{p}}_j$ given in Table 3.3 by A , we obtain

$$\sigma_j \underline{q}_j = U \underline{q}_{j-1} + \gamma_j \tilde{\underline{q}}_{j-1}, \quad \text{and} \quad \tilde{\underline{q}}_j = (\sigma_j \tilde{\underline{q}}_{j-1} + \bar{\gamma}_j \underline{q}_j).$$

Also, since $A \underline{p}_j = \rho U \underline{p}_j + \xi \underline{p}_j$, we can write

$$U \underline{p}_j = \frac{1}{\rho} (\underline{q}_j - \xi \underline{p}_j).$$

Using this information we rewrite the above minimal residual algorithm for shifted unitary matrices.

ALGORITHM 3.1 A Minimal Residual Algorithm

for Shifted Unitary Matrices:

$$\begin{aligned}\underline{p}_0 &= \tilde{\underline{p}}_0 = \underline{r}_0 / \|\underline{r}_0\|_{A^*A}, \\ \underline{q}_0 &= \tilde{\underline{q}}_0 = A \underline{p}_0, \\ \underline{x}_1 &= \underline{x}_0 + \alpha_0 \underline{p}_0, \quad \alpha_0 = \langle \underline{r}_0, \underline{q}_0 \rangle,\end{aligned}$$

$$\begin{aligned}
\underline{r}_1 &= \underline{r}_0 - \alpha_0 \underline{q}_0, \\
&\vdots \\
\gamma_j &= -\langle U \underline{q}_{j-1}, \tilde{\underline{q}}_{j-1} \rangle, \\
\sigma_j &= (1 - |\gamma_j|^2)^{1/2}, \\
\underline{q}_j &= \frac{1}{\sigma_j} (U \underline{q}_{j-1} + \gamma_j \tilde{\underline{q}}_{j-1}), \\
\underline{p}_j &= \frac{1}{\rho \sigma_j} (\underline{q}_{j-1} - \xi \underline{p}_{j-1}) + \frac{\gamma_j}{\sigma_j} \tilde{\underline{p}}_{j-1}, \\
\tilde{\underline{q}}_j &= \sigma_j \tilde{\underline{q}}_{j-1} + \bar{\gamma}_j \underline{q}_j, \\
\tilde{\underline{p}}_j &= \sigma_j \tilde{\underline{p}}_{j-1} + \bar{\gamma}_j \underline{p}_j, \\
\underline{x}_{j+1} &= \underline{x}_j + \alpha_j \underline{p}_j, \quad \alpha_j = \langle \underline{r}_j, \underline{q}_j \rangle, \\
\underline{r}_{j+1} &= \underline{r}_j - \alpha_j \underline{q}_j.
\end{aligned}$$

This algorithm requires the storage of 7 vectors:

$$\underline{x}_j, \underline{r}_j, \underline{p}_{j-1}, \tilde{\underline{p}}_{j-1}, \underline{q}_{j-1}, \tilde{\underline{q}}_{j-1}, U \underline{q}_{j-1}.$$

The approximate cost per iteration is:

- (1) 1 Matrix vector multiplication: $U \underline{q}_{j-1}$,
- (2) 2 inner products: $\langle \underline{r}_j, \underline{q}_j \rangle$, $\langle U \underline{q}_{j-1}, \tilde{\underline{q}}_{j-1} \rangle$,
- (3) 7 SAXPY.

The computational and storage requirements are comparable to the Jagels and Reichel algorithm (Table 3.2), which requires approximately 1 matrix vector multiplication, 1 inner product, 6 SAXPY, and the storage of 6 vectors.

Computationally, there is no advantage over the Jagels and Reichel algorithm. However, put in this context, it is clear that the direction vectors are A^*A -orthonormal. Thus, unitary, and shifted unitary matrices are examples of matrices that are not A^*A -normal(s), for some small degree s , but for which an A^*A -orthonormal basis can be constructed using a short multiple recursion. Therefore, the class of matrices for which an economical conjugate gradient algorithm exists is wider than that given by Faber and Manteuffel [5]. This motivates the study of alternate forms of recursions for constructing a B -orthogonal basis.

4. Multiple Recursion Formulation

4.1 Introduction

In Chapter 3, we saw that if a matrix A is either of the form

$$A = \rho U_B + \xi I, \quad \text{or} \quad A = U_B,$$

where U_B is B -unitary, then a double recursion,

$$\begin{aligned} \sigma_j \underline{p}_j &= U_B \underline{p}_{j-1} + \gamma_j \tilde{\underline{p}}_{j-1}, \\ \tilde{\underline{p}}_j &= \sigma_j \tilde{\underline{p}}_{j-1} + \bar{\gamma}_j \underline{p}_j, \end{aligned} \tag{4.1}$$

can be used to construct a B -orthonormal basis for $\mathcal{K}_d(\underline{r}_0, A)$. It was noted that a short $(s+2)$ -term recursion of the form given in (2.11) could not be used to construct this basis since these matrices are not B -normal(s) for some small degree s . That is, the B -adjoint of A cannot be written as

$$A^\dagger = p_s(A), \quad \text{for } s \text{ small.}$$

However, notice that if A is B -unitary, $A^\dagger = A^{-1}$. If λ_j is an eigenvalue of A , then $\frac{1}{\lambda_j}$ is an eigenvalue of A^\dagger , and we obtain,

$$A^\dagger = \frac{p_0(A)}{q_1(A)} = \frac{I}{A}, \quad \bar{\lambda}_j = \frac{p_0(\lambda_j)}{q_1(\lambda_j)},$$

where $p_0(\mu) = 1$ and $q_1(\mu) = \mu$. This motivates the exploration of a more general class of normal matrices, whose B -adjoint can be written as a ratio of two polynomials. We call this class B -normal(ℓ, m) matrices. They are defined and characterized in Section 4.2.

Consider the recursions in (4.1) for constructing a B -orthonormal basis. At step $j+1$, form

$$\sigma_{j+1} \underline{p}_{j+1} = U_B \underline{p}_j + \gamma_{j+1} \tilde{\underline{p}}_j.$$

For $\tilde{\underline{p}}_j$, substitute the second equation in (4.1) to obtain

$$\sigma_{j+1}\underline{p}_{j+1} = U_B\underline{p}_j + \gamma_{j+1}(\sigma_j\tilde{\underline{p}}_{j-1} + \bar{\gamma}_j\underline{p}_j).$$

Solving the first equation in (4.1) for $\tilde{\underline{p}}_{j-1}$, and making this substitution into the above yields the recursion,

$$\underline{p}_{j+1} = \frac{1}{\sigma_{j+1}}U_B\underline{p}_j - \frac{\gamma_{j+1}\sigma_j}{\gamma_j\sigma_{j+1}}U_B\underline{p}_{j-1} + \left(\frac{\gamma_{j+1}\sigma_j^2}{\gamma_j\sigma_{j+1}} + \frac{\gamma_{j+1}\bar{\gamma}_j}{\sigma_{j+1}}\right)\underline{p}_j. \quad (4.2)$$

Recall from Section 3.2, that the σ 's are nonzero normalization constants. For B -unitary and shifted B -unitary matrices, the recursions in (4.1) can be rewritten as the above single recursion as long as the γ 's are nonzero. When $A = \rho U_B + \xi I$, another way of writing the single recursion follows from substituting $U_B = \frac{1}{\rho}(A - \xi I)$, into (4.2) yielding,

$$\begin{aligned} \underline{p}_{j+1} &= \frac{1}{\rho\sigma_{j+1}}A\underline{p}_j - \frac{\gamma_{j+1}\sigma_j}{\rho\gamma_j\sigma_{j+1}}A\underline{p}_{j-1} \\ &+ \left(\frac{\gamma_{j+1}\sigma_j^2}{\gamma_j\sigma_{j+1}} - \frac{\xi}{\rho\sigma_{j+1}} + \frac{\gamma_{j+1}\bar{\gamma}_j}{\sigma_{j+1}}\right)\underline{p}_j + \frac{\gamma_{j+1}\sigma_j\xi}{\rho\gamma_j\sigma_{j+1}}\underline{p}_{j-1}. \end{aligned} \quad (4.3)$$

In Section 4.3, we begin by considering general recursions of the form,

$$\underline{p}_{j+1} = \sum_{i=j-t}^j \beta_{i,j}A\underline{p}_i - \sum_{i=j-s}^j \sigma_{i,j}\underline{p}_i,$$

where s and t can be any integers ≥ 0 . We will refer to this type of a recursion as a single (s, t) -recursion. The recursions given in (4.2) and (4.3) for B -unitary and shifted B -unitary matrices are of this type. By taking $t = 0$, we notice that the $(s + 2)$ -term recursion in (2.11) for B -normal(s) matrices is also of this form. We can view the single (s, t) -recursion as a more general form of short recursion, since it includes the short recurrences for B -normal(s) matrices, as well as those for B -unitary and shifted B -unitary matrices.

In the absence of a condition we call *breakdown*, which is analogous to the γ 's being zero in (4.2) and (4.3), this section demonstrates that a single (s, t) -recursion can be used to construct a B -orthogonal basis for $\mathcal{K}_d(\underline{x}_0, A)$ when A is B -normal(ℓ, m). This result is then extended to more general matrices which includes low rank perturbations of B -normal(ℓ, m) matrices. Since breakdown is possible in the single (s, t) -recursion, which is illustrated by an example given in Section 4.5, we show how the computations can be reorganized as a set of multiple recursions that avoids the problem of breakdown. Sufficient conditions on the system matrix A are then given in order for a B -orthogonal basis to be constructed using this form of multiple recursion.

We conclude this chapter with a brief summary, along with a few words about the likelihood of a breakdown occurring in the single (s, t) -recursion.

4.2 B -Normal(ℓ, m) Matrices

The unitary matrix was given as an example in Section 4.1 as a normal matrix whose adjoint can be expressed as the ratio of two polynomials,

$$A^* = \frac{p_0(A)}{q_1(A)} = \frac{I}{A}.$$

This section considers a general class of normal matrices of this type, called B -normal(ℓ, m) matrices.

DEFINITION 4.1 A is B -normal(ℓ, m) if A is B -normal and there exists polynomials $p_\ell(\lambda)$ and $q_m(\lambda)$, of degree ℓ and m respectively, such that

$$A^\dagger q_m(A) = p_\ell(A). \tag{4.4}$$

Since $A^\dagger = \frac{p_\ell(A)}{q_m(A)}$, we say A is B -normal with rational degree ℓ/m .

Notice that if $m = 0$, (4.4) becomes

$$A^\dagger = \hat{p}_\ell(A),$$

and A is B -normal(ℓ).

Next we characterize B -normal(ℓ, m) matrices. B -normal(ℓ, m) matrices satisfy

$$A^\dagger = \frac{p_\ell(A)}{q_m(A)} = \frac{a_\ell A^\ell + \cdots + a_1 A + a_0 I}{b_m A^m + \cdots + b_1 A + b_0 I}, \quad (4.5)$$

for some polynomials,

$$p_\ell(\mu) = a_\ell \mu^\ell + \cdots + a_1 \mu + a_0 \quad \text{and} \quad q_m(\mu) = b_m \mu^m + \cdots + b_1 + b_0,$$

of degree ℓ and m respectively. In the following Theorem, which characterizes B -normal(ℓ, m) matrices, we will assume that ℓ and m are the smallest degrees for which (4.5) holds. This means that $a_\ell, b_m \neq 0$, and that $p_\ell(\mu)$ and $q_m(\mu)$ have no common roots, since otherwise, (4.5) would hold for some smaller degrees ℓ' and m' .

THEOREM 4.1 Let A be B -normal(ℓ, m), where ℓ and m are the smallest degrees for which

$$A^\dagger = \frac{p_\ell(A)}{q_m(A)}.$$

Let $d(A)$ be the degree of the minimal polynomial of A . Then,

- (1) If $\ell > m + 1$, then $d(A) \leq \ell^2$,
- (2) if $\ell = m + 1$, then $d(A) \leq \ell^2$ or A is B -normal($1, 0$),
- (3) if $\ell < m$, and $b_0 \neq 0$, then $d(A) \leq m^2 + 1$,
- (4) if $\ell < m - 1$ and $b_0 = 0$, then $d(A) \leq m^2$,
- (5) if $\ell = m - 1$ and $b_0 = 0$, then $d(A) \leq m^2$, or A is B -normal($0, 1$),
- (6) if $\ell = m$, then $d(A) \leq m^2 + 1$, or A is B -normal($1, 1$).

Proof: From (4.5) it follows that the eigenvalues of a B -normal(ℓ, m) matrix can be written as:

$$\bar{\lambda}_i = \frac{p_\ell(\lambda_i)}{q_m(\lambda_i)}, \quad \forall \lambda_i \in \Sigma(A). \quad (4.6)$$

We assume that $p_\ell(\lambda_i) \neq 0$, $\forall \lambda_i \in \Sigma(A)$, since otherwise, (4.6) would imply that $\lambda_i = 0$ and A is singular. We further assume that $q_m(\lambda_i) \neq 0$, $\forall \lambda_i \in \Sigma(A)$, since $q_m(\lambda_i) = 0$ and $p_\ell(\lambda_i) \neq 0$ for some $\lambda_i \in \Sigma(A)$ would imply that $\lambda_i = \infty$. Taking conjugates in (4.6) yields

$$\lambda_i = \frac{\bar{p}_\ell(\bar{\lambda}_i)}{\bar{q}_m(\bar{\lambda}_i)} = \frac{\bar{p}_\ell\left(\frac{p_\ell(\lambda_i)}{q_m(\lambda_i)}\right)}{\bar{q}_m\left(\frac{p_\ell(\lambda_i)}{q_m(\lambda_i)}\right)}, \quad \forall \lambda_i \in \Sigma(A).$$

This can be expanded into

$$\begin{aligned} \lambda_i &= \frac{\bar{a}_\ell\left(\frac{p_\ell(\lambda_i)}{q_m(\lambda_i)}\right)^\ell + \bar{a}_{\ell-1}\left(\frac{p_\ell(\lambda_i)}{q_m(\lambda_i)}\right)^{\ell-1} + \cdots + \bar{a}_0}{\bar{b}_m\left(\frac{p_\ell(\lambda_i)}{q_m(\lambda_i)}\right)^m + \bar{b}_{m-1}\left(\frac{p_\ell(\lambda_i)}{q_m(\lambda_i)}\right)^{m-1} + \cdots + \bar{b}_0} \\ &= \frac{(q_m(\lambda_i))^m}{(q_m(\lambda_i))^\ell} \frac{\bar{a}_\ell(p_\ell(\lambda_i))^\ell + \bar{a}_{\ell-1}(p_\ell(\lambda_i))^{\ell-1}q_m(\lambda_i) + \cdots + \bar{a}_0(q_m(\lambda_i))^\ell}{\bar{b}_m(p_\ell(\lambda_i))^m + \bar{b}_{m-1}(p_\ell(\lambda_i))^{m-1}q_m(\lambda_i) + \cdots + \bar{b}_0(q_m(\lambda_i))^m}, \end{aligned}$$

$\forall \lambda_i \in \Sigma(A)$. Cross multiplication and collection of terms on one side yields

$$\begin{aligned} \lambda_i(q_m(\lambda_i))^\ell [\bar{b}_m(p_\ell(\lambda_i))^m + \bar{b}_{m-1}(p_\ell(\lambda_i))^{m-1}q_m(\lambda_i) + \cdots + \bar{b}_0(q_m(\lambda_i))^m] \\ - (q_m(\lambda_i))^m [\bar{a}_\ell(p_\ell(\lambda_i))^\ell + \bar{a}_{\ell-1}(p_\ell(\lambda_i))^{\ell-1}q_m(\lambda_i) + \cdots + \bar{a}_0(q_m(\lambda_i))^\ell] = 0, \end{aligned} \quad (4.7)$$

$\forall \lambda_i \in \Sigma(A)$.

First, consider the case when A is B -normal(ℓ, m) with $\ell > m$. Since $q_m(\lambda_i) \neq 0$, $\forall \lambda_i \in \Sigma(A)$, we can divide (4.7) by $(q_m(\lambda_i))^m$. It follows that the polynomial

$$\begin{aligned} \mu(q_m(\mu))^{\ell-m} [\bar{b}_m(p_\ell(\mu))^m + \bar{b}_{m-1}(p_\ell(\mu))^{m-1}q_m(\mu) + \cdots + \bar{b}_0(q_m(\mu))^m] \\ - \bar{a}_\ell(p_\ell(\mu))^\ell - \bar{a}_{\ell-1}(p_\ell(\mu))^{\ell-1}q_m(\mu) - \cdots - \bar{a}_0(q_m(\mu))^\ell = 0, \end{aligned} \quad (4.8)$$

whenever $\mu = \lambda_i$, and $\lambda_i \in \Sigma(A)$.

Proof of 1): If $\ell > m + 1$ and $m = 0$, A is B -normal(ℓ). In [5] it was shown that these matrices have less than or equal to ℓ^2 distinct eigenvalues. Therefore, $d(A) \leq \ell^2$.

If $\ell > m + 1$ and $m > 0$, the highest degree term of the polynomial in (4.8) is

$$-\bar{a}_\ell(a_\ell\mu^\ell)^\ell,$$

with degree ℓ^2 . By hypothesis, $a_\ell \neq 0$, which implies (4.8) has at most ℓ^2 distinct roots. Thus, $d(A) \leq \ell^2$.

Proof of 2): If $\ell = m + 1$ and $m = 0$, A is B -normal(1). It was shown in [5] that the eigenvalues of these matrices are collinear, that is, they lie on a straight line in the complex plane.

If $\ell = m + 1$ and $m > 0$, the highest degree term of the polynomial in (4.8) is

$$\mu(b_m\mu^m)\bar{b}_m(a_\ell\mu^\ell)^m - \bar{a}_\ell(a_\ell\mu^\ell)^\ell,$$

with degree ℓ^2 . Either the polynomial has at most ℓ^2 distinct roots, or (4.8) holds for all μ . In particular, (4.8) holds for the roots of $q_m(\mu)$, say μ_j , for $j = 1, \dots, m$. Plugging μ_j into (4.8) yields

$$\bar{a}_\ell(p_\ell(\mu_j))^\ell = 0, \quad \text{for } j = 1, \dots, m.$$

This means that either $a_\ell = 0$ or $p_\ell(\mu_j) = 0$, for $j = 1, \dots, m$, or both are zero. By hypothesis, $a_\ell \neq 0$ and $p_\ell(\mu)$ and $q_m(\mu)$ have no common roots, so it follows that (4.8) cannot hold for all μ , thus $d(A) \leq \ell^2$.

Next, consider the case when A is B -normal(ℓ, m), with $\ell \leq m$. Since $q_m(\lambda_i) \neq 0$, $\forall \lambda_i \in \Sigma(A)$, we can divide (4.7) by $(q_m(\lambda_i))^\ell$. We obtain the polynomial

$$\begin{aligned} & \mu [\bar{b}_m(p_\ell(\mu))^m + \bar{b}_{m-1}(p_\ell(\mu))^{m-1}q_m(\mu) + \dots + \bar{b}_0(q_m(\mu))^m] \\ & - (q_m(\mu))^{m-\ell} [\bar{a}_\ell(p_\ell(\mu))^\ell + \bar{a}_{\ell-1}(p_\ell(\mu))^{\ell-1}q_m(\mu) + \dots + \bar{a}_0(q_m(\mu))^\ell] = 0, \end{aligned} \quad (4.9)$$

whenever $\mu = \lambda_i$, and $\lambda_i \in \Sigma(A)$. Since $p_\ell(\lambda)$ and $q_m(\lambda)$ are polynomials of degree ℓ and m respectively, a_ℓ and b_m are nonzero. However, it is possible that any of the other coefficients, $a_0, \dots, a_{\ell-1}$ and b_0, \dots, b_{m-1} , could be zero.

Proof of 3): If $\ell < m$ and $b_0 \neq 0$, the highest degree term of the polynomial in (4.9) is

$$\bar{b}_0 \mu (b_m \mu^m)^m,$$

with degree $m^2 + 1$. By hypothesis, $b_0, b_m \neq 0$, and it follows that $d(A) \leq m^2 + 1$.

Proof of 4): Suppose that $\ell < m - 1$ and $b_0 = 0$. We assume that $a_0 \neq 0$, for otherwise we could factor μ out of both $p_\ell(\mu)$ and $q_m(\mu)$, showing that A is really B -normal($\ell - 1, m - 1$). The highest degree term of the polynomial in (4.9) is

$$\bar{a}_0 (b_m \mu^m)^{m-\ell} (b_m \mu^m)^\ell = \bar{a}_0 (b_m \mu^m)^m,$$

with degree m^2 . By hypothesis, $a_0, b_m \neq 0$, and it follows that (4.9) cannot be zero for all μ , thus, $d(A) \leq m^2$.

Proof of 5): Suppose that $\ell = m - 1$ and $b_0 = 0$. We assume that $a_0 \neq 0$, since otherwise this would imply that A is B -normal($\ell - 1, m - 1$). There are 2 possibilities to consider, $\ell > 0$ and $\ell = 0$.

If $\ell > 0$, then $m > 1$ and it is possible for b_1 to be zero. If $b_1 = 0$, the highest degree term of the polynomial in (4.9) is

$$\bar{a}_0 (b_m \mu^m)^{m-\ell} (b_m \mu^m)^\ell = \bar{a}_0 (b_m \mu^m)^m,$$

with degree m^2 . Since a_0 and b_m are nonzero by hypothesis, it follows that the polynomial has at most m^2 distinct roots, thus $d(A) \leq m^2$. If $b_1 \neq 0$, the highest degree term of the polynomial in (4.9) is

$$\bar{b}_1 \mu (b_m \mu^m)^{m-1} (a_\ell \mu^\ell) - \bar{a}_0 (b_m \mu^m)^m,$$

with degree m^2 . Again, either $d(A) \leq m^2$, or (4.9) holds for all μ . In particular, it must hold for the roots of $p_\ell(\mu)$, say μ_j , for $j = 1, \dots, \ell$. Plugging μ_j , for $j = 1, \dots, \ell$ into (4.9) yields

$$\bar{a}_0 (q_m(\mu_j))^m = 0, \quad j = 1, \dots, \ell.$$

To satisfy (4.9) for all μ , requires that either $a_0 = 0$, or $q_m(\mu_j) = 0$, for $j = 1, \dots, \ell$, or both are zero. By hypothesis, $a_0 \neq 0$, and $p_\ell(\mu)$ and $q_m(\mu)$ have no common roots, thus, $d(A) \leq m^2$.

If $\ell = 0$, then $m = 1$, and it follows by hypothesis that $b_m = b_1 \neq 0$, $a_\ell = a_0 \neq 0$, and $b_0 = 0$. These matrices are B -normal(0, 1), where in addition, $b_0 = 0$. Therefore, $p_0(\mu) = a_0$ and $q_m(\mu) = b_1\mu$, and it follows that (4.6) becomes

$$\bar{\lambda}_i = \frac{a_0}{b_1\lambda_i}, \quad \forall \lambda_i \in \Sigma(A),$$

or,

$$\bar{\lambda}_i\lambda_i = \frac{a_0}{b_1}, \quad \forall \lambda_i \in \Sigma(A).$$

Therefore, $\frac{a_0}{b_1}$ must be positive and real. Matrices that are B -normal(0, 1), with $b_0 = 0$, have eigenvalues that lie on a circle of radius $\sqrt{\frac{a_0}{b_1}}$. That is, they are scaled unitary matrices.

Proof of 6): When A is B -normal(ℓ, m) with $\ell = m$, (4.5) becomes

$$A^\dagger = \frac{p_m(A)}{q_m(A)} = \frac{a_m}{b_m} + \frac{\tilde{p}_{m-1}(A)}{q_m(A)},$$

where the second equality follows upon division of $p_m(A)$ by $q_m(A)$. Denoting $\hat{A} = (A - \zeta I)$, where $\zeta = \frac{\tilde{a}_m}{b_m}$, the above can be rearranged yielding

$$\hat{A}^\dagger = \frac{\tilde{p}_{m-1}(A)}{q_m(A)} = \frac{\hat{p}_{m-1}(\hat{A})}{\hat{q}_m(\hat{A})}.$$

This shows that if A is B -normal(m, m), then there exists a constant ζ , such that $A - \zeta I$ is B -normal($m - 1, m$). We apply the results from the $\ell < m$ case to the matrix \hat{A} . It follows that if A is B -normal(m, m), the only cases that yield more than $m^2 + 1$ distinct eigenvalues are B -normal(1, 1) matrices whose eigenvalues can be obtained by shifting the eigenvalues of a scaled unitary matrix by some constant ζ . These are the shifted unitary matrices described in Section 2. \square

The above Theorem shows that if A is B -normal(ℓ, m) and either ℓ or m are greater than one, then A has a relatively small number of distinct eigenvalues. Furthermore, from the proof, we see that if $\ell, m \leq 1$, the only matrices that have more than 2 distinct eigenvalues are B -normal(1), B -unitary, and shifted B -unitary matrices.

4.3 A Multiple Recursion For B -Normal(ℓ, m) Matrices

In this section, we will begin by considering general recursions of the form,

$$\underline{p}_{j+1} = \sum_{k=j-t}^j \beta_{k,j} A \underline{p}_k - \sum_{i=j-s}^j \sigma_{i,j} \underline{p}_i,$$

where s and t can be any integers ≥ 0 . We refer to this as a single (s, t) -recursion.

Related work involving alternate forms of short recursions can be found in ([3], [4], [11], and [13]). The work in ([3] and [4]) is based upon a generalized Krylov subspace. The vectors that span this subspace involve powers of both A and A^* . The methods studied in [13] differ from the work in this thesis in that these methods are not conjugate gradient-like methods (see Chapter 2). The work in [11] relates to the work in this thesis because conjugate gradient methods can be viewed as operator coefficient methods.

At each step in the (s, t) -iteration, the term $\beta_{j,j} A \underline{p}_j$ brings the recursion up to the next higher dimension Krylov subspace yielding $\underline{p}_{j+1} \in \mathcal{K}_{j+2}(\underline{x}_0, A)$. In order for the above recursion to yield a B -orthogonal basis, $\beta_{j,j} \neq 0$ for every j . Since $\beta_{j,j}$ is usually a normalization constant, for simplicity of this presentation, we will assume $\beta_{j,j} = 1$ for every j , and denote the single (s, t) -recursion as,

$$\underline{p}_{j+1} = A \underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} A \underline{p}_k - \sum_{i=j-s}^j \sigma_{i,j} \underline{p}_i, \quad (4.10)$$

where s and t can be any integers ≥ 0 . Notice that, \underline{p}_{j+1} can always be computed

by a recursion of the form

$$\underline{p}_{j+1} = A\underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} A\underline{p}_k - \sum_{i=0}^j \sigma_{i,j} \underline{p}_i,$$

because, given any $\{\beta_{k,j}\}_{k=j-t}^{j-1}$, \underline{p}_{j+1} can be constructed so it is B -orthogonal to $\text{sp}\{\underline{p}_0, \dots, \underline{p}_j\}$ by choosing

$$\sigma_{i,j} = \frac{\langle A\underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} A\underline{p}_k, \underline{p}_i \rangle_B}{\langle \underline{p}_j, \underline{p}_i \rangle_B}, \quad i = 0, \dots, j.$$

However, we are interested in determining when a B -orthogonal basis for $\mathcal{K}_d(\underline{r}_0, A)$ can be constructed using a short recursion of the form (4.10), that is, with both s and t small. Notice that in order for this to be the case, at each step, $j+1$, $\sigma_{i,j} = 0$, for $i = 0, \dots, j-s-1$.

To be more precise, we say that a B -orthogonal basis for $\mathcal{K}_d(\underline{r}_0, A)$ can be constructed using the single (s, t) -recursion (4.10) if for every \underline{p}_0 and every $0 \leq j \leq d(\underline{p}_0, A) - 2$, there exists coefficients $\{\beta_{k,j}\}_{k=j-t}^{j-1}$, such that

$$\sigma_{i,j} = \frac{\langle A\underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} A\underline{p}_k, \underline{p}_i \rangle_B}{\langle \underline{p}_j, \underline{p}_i \rangle_B} = 0, \quad \text{for } i = 0, \dots, j-s-1,$$

or equivalently,

$$\langle \underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} \underline{p}_k, A^\dagger \underline{p}_i \rangle_B = 0, \quad \text{for } i = 0, \dots, j-s-1. \quad (4.11)$$

Suppose A is B -normal (ℓ, m) . According to Definition 4.1, there exists polynomials $p_\ell(\mu)$ and $q_m(\mu)$, of degrees ℓ and m respectively, such that

$$A^\dagger = \frac{p_\ell(A)}{q_m(A)}.$$

Since for every i , $\underline{p}_i \in \mathcal{K}_{i+1}(\underline{p}_0, A)$, \underline{p}_i can be written as $\psi_i(A)\underline{p}_0$, for some polynomial ψ_i of exact degree i . There exists polynomials, $\tilde{\psi}_{i-m}$, and $r_{m-1}^{(i)}$ such that

$$\underline{p}_i = \psi_i(A)\underline{p}_0 = q_m(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 + r_{m-1}^{(i)}(A)\underline{p}_0,$$

where the second equality follows upon division of $\psi_i(\mu)$ by $q_m(\mu)$, resulting in a quotient term, $\tilde{\psi}_{i-m}(A)$, and a remainder term, $r_{m-1}^{(i)}(A)$. Since by hypothesis, $A^\dagger q_m(A) = p_\ell(A)$, it follows that

$$\begin{aligned} A^\dagger \underline{p}_i &= A^\dagger q_m(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 + A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0 \\ &= p_\ell(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 + A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0. \end{aligned} \quad (4.12)$$

Recall that in order for an (s, t) -recursion of the form (4.10) to yield a B -orthogonal basis, at each step $j+1$, there must exist coefficients $\{\beta_{k,j}\}_{k=j-t}^{j-1}$ satisfying (4.11). By letting $(s, t) = (\ell, m)$ in (4.11), and then substituting in the expression for $A^\dagger \underline{p}_i$ given in (4.12), it follows that for $i = 0, \dots, j - \ell - 1$, we must satisfy

$$\langle \underline{p}_j + \sum_{k=j-m}^{j-1} \beta_{k,j} \underline{p}_k, p_\ell(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 + A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0 \rangle_B = 0.$$

Since $r_{m-1}^{(i)}(A)$ is a polynomial of degree $\leq m-1$,

$$r_{m-1}^{(i)}(A) \underline{p}_0 \in \text{sp}\{\underline{p}_0, A \underline{p}_0, \dots, A^{m-1} \underline{p}_0\} = \text{sp}\{\underline{p}_0, \dots, \underline{p}_{m-1}\},$$

and thus,

$$A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0 \in \text{sp}\{A^\dagger \underline{p}_0, A^\dagger \underline{p}_1, \dots, A^\dagger \underline{p}_{m-1}\}.$$

For $i = 0, \dots, j - \ell - 1$,

$$p_\ell(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 \in \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{j-m-1}\}.$$

By orthogonality, $\text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{j-m-1}\}$ is B -orthogonal to $\text{sp}\{\underline{p}_{j-m}, \dots, \underline{p}_j\}$, and it follows that we only need to choose $\{\beta_{k,j}\}_{k=j-m}^{j-1}$ so that

$$\langle \underline{p}_j + \sum_{k=j-m}^{j-1} \beta_{k,j} \underline{p}_k, A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0 \rangle_B = 0, \quad \text{for } i = 0, \dots, j - \ell - 1.$$

This can be accomplished by choosing $\{\beta_{k,j}\}_{k=j-m}^{j-1}$ to satisfy

$$\begin{bmatrix} \langle \underline{p}_{j-m}, A^\dagger \underline{p}_0 \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_0 \rangle_B & \langle \underline{p}_j, A^\dagger \underline{p}_0 \rangle_B \\ \langle \underline{p}_{j-m}, A^\dagger \underline{p}_1 \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_1 \rangle_B & \langle \underline{p}_j, A^\dagger \underline{p}_1 \rangle_B \\ \vdots & & \vdots & \vdots \\ \langle \underline{p}_{j-m}, A^\dagger \underline{p}_{m-1} \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_{m-1} \rangle_B & \langle \underline{p}_{i-m}, A^\dagger \underline{p}_{m-1} \rangle_B \end{bmatrix} \begin{pmatrix} \beta_{j-m,j} \\ \vdots \\ \beta_{j-1,j} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix},$$

or equivalently, to satisfy

$$\begin{bmatrix} \langle \underline{p}_{j-m}, A^\dagger \underline{p}_0 \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_0 \rangle_B \\ \langle \underline{p}_{j-m}, A^\dagger \underline{p}_1 \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_1 \rangle_B \\ \vdots & & \vdots \\ \langle \underline{p}_{j-m}, A^\dagger \underline{p}_{m-1} \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_{m-1} \rangle_B \end{bmatrix} \begin{pmatrix} \beta_{j-m,j} \\ \vdots \\ \beta_{j-1,j} \end{pmatrix} = - \begin{pmatrix} \langle \underline{p}_j, A^\dagger \underline{p}_0 \rangle_B \\ \langle \underline{p}_j, A^\dagger \underline{p}_1 \rangle_B \\ \vdots \\ \langle \underline{p}_j, A^\dagger \underline{p}_{m-1} \rangle_B \end{pmatrix}. \quad (4.13)$$

If the above system is consistent at every step, then an (s, t) -recursion, with $(s, t) = (\ell, m)$ can be used to construct a B -orthogonal basis for B -normal (ℓ, m) matrices.

We say that a breakdown occurs in the single (s, t) -recursion if the above system becomes inconsistent at some step in the iteration. If this happens, the matrix in (4.13) is singular. Recall the matrix equation (2.10) obtained after a given step in the iteration using a full Gram-Schmidt process to construct a B -orthogonal basis. After constructing \underline{p}_j , this is given by,

$$AP_j = P_{j+1}H_{j+1,j}.$$

Notice that the matrix in (4.13) corresponds to the upper right $m \times m$ corner of the Hessenberg matrix $H_{j+1,j}$. We can correlate breakdown in the single (s, t) -recursion to the singularity of a minor of the Hessenberg matrix. We will discuss the likelihood of these matrices being singular in Chapter 6, since this will become important later in our analysis.

For now, we note that in the absence of breakdown, an (s, t) -recursion, with $(s, t) = (\ell, m)$, can be applied to any B -normal (ℓ, m) matrix. However, in order for this construction to be economical, ℓ and m must not be too large.

Next, we will show how to reformulate this iteration in terms of multiple recursions to avoid the problem of breakdown. Suppose $d(\underline{p}_0, A) = N$, and recall that for any matrix A , a B -orthogonal basis can always be computed with a full recursion given by (2.8). After N steps we obtain the corresponding matrix equation (2.9),

where H_N is an upper Hessenberg matrix, with entries:

$$h_{i,j} = \begin{cases} \sigma_{i,j} = \frac{\langle \underline{p}_j, A^\dagger \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B}, & \text{if } j \geq i \\ 1, & \text{if } j = i - 1 \\ 0, & \text{if } j < i - 1 \end{cases} \quad (4.14)$$

Consider the upper triangular part of H_N . Substituting the expression for $A^\dagger \underline{p}_i$ given in (4.12) into $\sigma_{i,j}$ yields

$$\sigma_{i,j} = \frac{\langle \underline{p}_j, p_\ell(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 + A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0 \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B}, \quad \text{for } j \geq i.$$

Note that

$$p_\ell(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 \in \text{sp}\{\underline{p}_0, A \underline{p}_0, \dots, A^{i-m+\ell} \underline{p}_0\} = \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{i-m+\ell}\},$$

so that for $j > i - m + \ell$, $\underline{p}_j \perp_B p_\ell(A) \tilde{\psi}_{i-m}(A) \underline{p}_0$, and

$$\sigma_{i,j} = \frac{\langle \underline{p}_j, A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0 \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B} = \frac{\langle A \underline{p}_j, r_{m-1}^{(i)}(A) \underline{p}_0 \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B}.$$

Since $r_{m-1}^{(i)}(A) \underline{p}_0 \in \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{m-1}\}$, we can write

$$r_{m-1}^{(i)}(A) \underline{p}_0 = (\underline{\rho}_i)_{m-1} \underline{p}_{m-1} + \dots + (\underline{\rho}_i)_0 \underline{p}_0,$$

for some vector of coefficients,

$$\underline{\rho}_i = [(\underline{\rho}_i)_0, \dots, (\underline{\rho}_i)_{m-1}]^T \in \mathcal{C}^m. \quad (4.15)$$

Denote

$$\underline{\eta}_j = [\langle A \underline{p}_j, \underline{p}_0 \rangle_B, \langle A \underline{p}_j, \underline{p}_1 \rangle_B, \dots, \langle A \underline{p}_j, \underline{p}_{m-1} \rangle_B]^T \in \mathcal{C}^m, \quad (4.16)$$

and note that

$$\begin{aligned} \langle A \underline{p}_j, r_{m-1}^{(i)}(A) \underline{p}_0 \rangle_B &= (\underline{\rho}_i)_{m-1} \langle A \underline{p}_j, \underline{p}_{m-1} \rangle_B + \dots + (\underline{\rho}_i)_0 \langle A \underline{p}_j, \underline{p}_0 \rangle_B \\ &= \langle \underline{\eta}_j, \underline{\rho}_i \rangle. \end{aligned}$$

It follows that the entries of H_N simplify, yielding:

$$h_{i,j} = \begin{cases} \frac{\langle \eta_j, \rho_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j > \max\{i-1, i-m+\ell\} \\ \frac{\langle A\underline{p}_j, \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j = i, \dots, i-m+\ell \quad (\text{if } \ell \geq m) \\ 1 & j = i-1 \\ 0 & j < i-1 \end{cases}.$$

Consider the following decomposition of the Hessenberg matrix H_N :

$$H_N = T + U,$$

where U is an $N \times N$ upper triangular matrix whose entries are given by:

$$u_{i,j} = \begin{cases} \frac{\langle \eta_j, \rho_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j \geq i \\ 0 & j < i \end{cases},$$

and T is $N \times N$ upper Hessenberg with entries given by:

$$t_{i,j} = \begin{cases} 0 & j > \max\{i-1, i-m+\ell\} \\ \frac{\langle A\underline{p}_j, \underline{p}_i \rangle_B - \langle \eta_j, \rho_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j = i, \dots, i-m+\ell \quad (\text{if } \ell \geq m) \\ 1 & j = i-1 \\ 0 & j < i-1 \end{cases}.$$

Notice that T has an upper bandwidth of $\max\{0, \ell - m + 1\}$, so that if $m > \ell$, T has only a nonzero subdiagonal consisting of all ones. This way of decomposing H_N yields a banded Hessenberg matrix T , and an upper triangular matrix U , that can be factored as follows:

$$U = \begin{bmatrix} \frac{\langle \eta_0, \rho_0 \rangle}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \frac{\langle \eta_1, \rho_0 \rangle}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \cdots & \frac{\langle \eta_{N-1}, \rho_0 \rangle}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ & \frac{\langle \eta_1, \rho_1 \rangle}{\langle \underline{p}_1, \underline{p}_1 \rangle_B} & \cdots & \frac{\langle \eta_{N-1}, \rho_1 \rangle}{\langle \underline{p}_1, \underline{p}_1 \rangle_B} \\ & & \ddots & \vdots \\ & & & \frac{\langle \eta_{N-1}, \rho_{N-1} \rangle}{\langle \underline{p}_{N-1}, \underline{p}_{N-1} \rangle_B} \end{bmatrix}$$

$$\begin{aligned}
&= \begin{bmatrix} \frac{\underline{\bar{p}}_0^T}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \frac{\underline{\bar{p}}_0^T}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \cdots & \frac{\underline{\bar{p}}_0^T}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ & \frac{\underline{\bar{p}}_1^T}{\langle \underline{p}_1, \underline{p}_1 \rangle_B} & \cdots & \frac{\underline{\bar{p}}_1^T}{\langle \underline{p}_1, \underline{p}_1 \rangle_B} \\ & & \ddots & \vdots \\ & & & \frac{\underline{\bar{p}}_{N-1}^T}{\langle \underline{p}_{N-1}, \underline{p}_{N-1} \rangle_B} \end{bmatrix} \begin{bmatrix} \underline{\eta}_0 & & & \\ & \underline{\eta}_1 & & \\ & & \ddots & \\ & & & \underline{\eta}_{N-1} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\underline{\bar{p}}_0^T}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & & & \\ & \ddots & & \\ & & \frac{\underline{\bar{p}}_{N-1}^T}{\langle \underline{p}_{N-1}, \underline{p}_{N-1} \rangle_B} & \\ & & & \end{bmatrix} \begin{bmatrix} I_m & \cdots & I_m \\ & \ddots & \vdots \\ & & I_m \end{bmatrix} \begin{bmatrix} \underline{\eta}_0 & & & \\ & \ddots & & \\ & & & \underline{\eta}_{N-1} \end{bmatrix} \\
&= M E D.
\end{aligned}$$

The dimensions of the above matrices are given by:

$$\begin{aligned}
M &\text{ is } N \times (N \times m), \\
E &\text{ is } (N \times m) \times (N \times m), \text{ and} \\
D &\text{ is } (N \times m) \times N.
\end{aligned}$$

The matrix E consists of the blocks I_m , on and above the diagonal, where each I_m is an $m \times m$ identity matrix. Combining this information, we obtain

$$\begin{aligned}
AP_N &= P_N H_N \\
&= P_N T + P_N U \\
&= P_N T + P_N M E D.
\end{aligned} \tag{4.17}$$

Define

$$Q = P_N M E. \tag{4.18}$$

Notice that Q is an $N \times (N \times m)$ matrix. In particular, we denote

$$Q = \left[\begin{array}{c|ccc|ccc|ccc} | & & | & | & & | & & | & & | \\ \underline{q}_{0_0} & \cdots & \underline{q}_{0_{m-1}} & \underline{q}_{1_0} & \cdots & \underline{q}_{1_{m-1}} & \cdots & \underline{q}_{N-1_0} & \cdots & \underline{q}_{N-1_{m-1}} \\ | & & | & | & & | & & | & & | \end{array} \right].$$

Substituting Q into (4.17) yields,

$$AP_N = P_N T + QD.$$

The following Theorem summarizes these results.

THEOREM 4.2 If A is B -normal(ℓ, m), then for every \underline{r}_0 , a B -orthogonal basis for $\mathcal{K}_d(\underline{r}_0, A)$ can be constructed using the multiple recursion given in (4.21).

Proof: See above discussion. \square

Recall that if breakdown in the single (s, t) -recursion occurs at step $j + 1$, the matrix given in (4.13) is singular. Using (4.16), this matrix can be rewritten as:

$$\left[\begin{array}{c|ccc|c} & & & & \\ & & & & \\ \hline & \eta_{j-m} & \cdots & \eta_{j-1} & \\ & & & & \\ \hline & & & & \end{array} \right]. \quad (4.22)$$

4.4 A Multiple Recursion For Generalizations Of B -Normal(ℓ, m) Matrices

Next, we will see that a short single (s, t) -recursion and a similar set of multiple recursions can actually be applied to more general matrices. When A is B -normal(ℓ, m), there exists polynomials $p_\ell(\mu)$ and $q_m(\mu)$ satisfying

$$A^\dagger q_m(A) - p_\ell(A) = 0.$$

Suppose instead that these polynomials satisfy

$$A^\dagger q_m(A) - p_\ell(A) = Q_B(A), \quad \text{where, } \text{Rank}(Q_B(A)) = \kappa. \quad (4.23)$$

We refer to a matrix that satisfies this relationship as a generalization of a B -normal(ℓ, m) matrix.

We first consider the construction of a B -orthogonal basis using a single (s, t) -recursion (4.10) for matrices satisfying (4.23). Recall that if a B -orthogonal basis can be constructed using a single (s, t) -recursion, then (4.11) must be satisfied at each step. Analogous to the B -normal(ℓ, m) case, there exists polynomials, ψ_i , $\tilde{\psi}_{i-m}$, and $r_{m-1}^{(i)}$ such that

$$\underline{p}_i = \psi_i(A)\underline{p}_0 = q_m(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 + r_{m-1}^{(i)}(A)\underline{p}_0,$$

where the second equality follows from dividing the polynomial $\psi_i(A)$ by $q_m(A)$. Multiplying through by A^\dagger and then adding and subtracting $p_\ell(A)\tilde{\psi}_{i-m}(A)\underline{p}_0$ yields

$$\begin{aligned} A^\dagger \underline{p}_i &= [A^\dagger q_m(A) - p_\ell(A)]\tilde{\psi}_{i-m}(A)\underline{p}_0 + p_\ell(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 + A^\dagger r_{m-1}^{(i)}(A)\underline{p}_0 \\ &= Q_B(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 + p_\ell(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 + A^\dagger r_{m-1}^{(i)}(A)\underline{p}_0. \end{aligned} \quad (4.24)$$

For an (s, t) -recursion to exist, at every step, there must exist coefficients $\{\beta_{k,j}\}_{k=j-t}^{j-1}$ satisfying (4.11). By substituting (4.24) into (4.11), it follows that for $i = 0, \dots, j - s - 1$, the β 's must satisfy

$$\langle \underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} \underline{p}_k, Q_B(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 + p_\ell(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 + A^\dagger r_{m-1}^{(i)}(A)\underline{p}_0 \rangle_B = 0.$$

Recall that $A^\dagger r_{m-1}^{(i)}(A)\underline{p}_0 \in \text{sp}\{A^\dagger \underline{p}_0, \dots, A^\dagger \underline{p}_{m-1}\}$, and let $\{\underline{\phi}_0, \underline{\phi}_1, \dots, \underline{\phi}_{\kappa-1}\}$ be a basis for the range of $Q_B(A)$. It follows that

$$Q_B(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 = (\underline{\tau}_i)_{\kappa-1} \underline{\phi}_{\kappa-1} + \dots + (\underline{\tau}_i)_0 \underline{\phi}_0,$$

for some vector of coefficients, $\underline{\tau}_i = [(\underline{\tau}_i)_0, \dots, (\underline{\tau}_i)_{\kappa-1}]^T \in \mathcal{C}^\kappa$.

By choosing s and t to satisfy

$$s - \ell \geq t - m \geq \kappa, \quad (4.25)$$

we have for $i = 0, \dots, j - s - 1$ that

$$\begin{aligned} i - m + \ell &\leq (j - s - 1) - m + \ell = j - 1 - m - (s - \ell) \\ &\leq j - 1 - m - (t - m) = j - 1 - t, \end{aligned}$$

and thus,

$$p_\ell(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 \in \text{sp}\{\underline{p}_0, \dots, \underline{p}_{i-m+\ell}\} \subseteq \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{j-t-1}\}.$$

Therefore,

$$p_\ell(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 -_B \left(\underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} \underline{p}_k \right),$$

and it follows that we only need to choose the $\beta_{k,j}$'s so that for $i = 0, \dots, j - s - 1$,

$$\langle \underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} \underline{p}_k, A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0 + Q_B(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 \rangle_B = 0.$$

This can be accomplished by choosing $\{\beta_{k,j}\}_{k=j-t}^{j-1}$ to satisfy

$$\begin{bmatrix} \langle \underline{p}_{j-t}, A^\dagger \underline{p}_0 \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_0 \rangle_B & \langle \underline{p}_j, A^\dagger \underline{p}_0 \rangle_B \\ \vdots & & \vdots & \vdots \\ \langle \underline{p}_{j-t}, A^\dagger \underline{p}_{m-1} \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_{m-1} \rangle_B & \langle \underline{p}_j, A^\dagger \underline{p}_{m-1} \rangle_B \\ \langle \underline{p}_{j-t}, \underline{\phi}_0 \rangle_B & \cdots & \langle \underline{p}_{j-1}, \underline{\phi}_0 \rangle_B & \langle \underline{p}_j, \underline{\phi}_0 \rangle_B \\ \vdots & & \vdots & \vdots \\ \langle \underline{p}_{j-t}, \underline{\phi}_{\kappa-1} \rangle_B & \cdots & \langle \underline{p}_{j-1}, \underline{\phi}_{\kappa-1} \rangle_B & \langle \underline{p}_j, \underline{\phi}_{\kappa-1} \rangle_B \end{bmatrix} \begin{pmatrix} \beta_{j-t,j} \\ \vdots \\ \beta_{j-1,t} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix},$$

or equivalently,

$$\begin{bmatrix} \langle \underline{p}_{j-t}, A^\dagger \underline{p}_0 \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_0 \rangle_B \\ \vdots & & \vdots \\ \langle \underline{p}_{j-t}, A^\dagger \underline{p}_{m-1} \rangle_B & \cdots & \langle \underline{p}_{j-1}, A^\dagger \underline{p}_{m-1} \rangle_B \\ \langle \underline{p}_{j-t}, \underline{\phi}_0 \rangle_B & \cdots & \langle \underline{p}_{j-1}, \underline{\phi}_0 \rangle_B \\ \vdots & & \vdots \\ \langle \underline{p}_{j-t}, \underline{\phi}_{\kappa-1} \rangle_B & \cdots & \langle \underline{p}_{j-1}, \underline{\phi}_{\kappa-1} \rangle_B \end{bmatrix} \begin{pmatrix} \beta_{j-t,j} \\ \vdots \\ \beta_{j-1,j} \end{pmatrix} = - \begin{pmatrix} \langle \underline{p}_j, A^\dagger \underline{p}_0 \rangle_B \\ \vdots \\ \langle \underline{p}_j, A^\dagger \underline{p}_{m-1} \rangle_B \\ \langle \underline{p}_j, \underline{\phi}_0 \rangle_B \\ \vdots \\ \langle \underline{p}_j, \underline{\phi}_{\kappa-1} \rangle_B \end{pmatrix}. \quad (4.26)$$

This system has $m + \kappa$ equations in t unknowns. From (4.25) we see that t was chosen so that $\kappa \leq t - m$, which means the system has $t - (m + \kappa)$ degrees of freedom. Up to this point, we have only specified that s and t must satisfy (4.25). There is some flexibility in choosing s and t . To make s and t as small as possible, yielding the most economical (s, t) -recursion, we would choose

$$t = \kappa + m, \quad \text{and} \quad s = \kappa + \ell,$$

so the system in (4.26) is $(m + \kappa) \times (m + \kappa)$. If this system is nonsingular for every step, then an (s, t) -recursion will yield a B -orthogonal basis. Notice that the number of terms s and t depends on the rank of $Q_B(A)$, as well as the degrees of the polynomials p_ℓ , and q_m .

Analogous to the B -normal(ℓ, m) case, we say that breakdown occurs in the single (s, t) -recursion if the above system becomes inconsistent at some step in the iteration. If this happens, the matrix in (4.26) is singular.

A derivation similar to that used in the B -normal(ℓ, m) case, yields another way of constructing a B -orthogonal basis for matrices of this form. This formulation also involves several short recursions at each step, and avoids the possibility of breakdown.

Suppose $d(\underline{p}_0, A) = N$. Recall that a B -orthogonal basis could be computed using a Gram-Schmidt process (2.8) which yields at step N the corresponding matrix equation (2.9). The Hessenberg matrix H_N has entries given by:

$$h_{i,j} = \begin{cases} \sigma_{i,j} = \frac{\langle \underline{p}_j, A^\dagger \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B}, & \text{if } j \geq i \\ 1, & \text{if } j = i - 1 \\ 0, & \text{if } j < i - 1 \end{cases} \quad (4.27)$$

Consider the upper triangular part of H_N . Substituting in the expression for $A^\dagger \underline{p}_i$ given in (4.24) yields,

$$\sigma_{i,j} = \frac{\langle \underline{p}_j, Q_B(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 + p_\ell(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 + A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0 \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B}.$$

Since $p_\ell(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 \in \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{i-m+\ell}\}$, for $j > i - m + \ell$,

$$\underline{p}_j -_B p_\ell(A) \tilde{\psi}_{i-m}(A) \underline{p}_0,$$

and,

$$\sigma_{i,j} = \frac{\langle \underline{p}_j, Q_B(A) \tilde{\psi}_{i-m}(A) \underline{p}_0 + A^\dagger r_{m-1}^{(i)}(A) \underline{p}_0 \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B}.$$

Since $r_{m-1}^{(i)}(A) \underline{p}_0 \in \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{m-1}\}$, we can write,

$$r_{m-1}^{(i)}(A) \underline{p}_0 = (\underline{\rho}_i)_{m-1} \underline{p}_{m-1} + \dots + (\underline{\rho}_i)_0 \underline{p}_0,$$

for some vector of coefficients,

$$\underline{\rho}_i = [(\underline{\rho}_i)_0, \dots, (\underline{\rho}_i)_{m-1}]^T \in \mathcal{C}^m. \quad (4.28)$$

Next, by denoting

$$\underline{\eta}_j = [\langle A\underline{p}_j, \underline{p}_0 \rangle_B, \langle A\underline{p}_j, \underline{p}_1 \rangle_B, \dots, \langle A\underline{p}_j, \underline{p}_{m-1} \rangle_B]^T \in \mathcal{C}^m, \quad (4.29)$$

we obtain

$$\begin{aligned} \langle \underline{p}_j, A^\dagger r_{m-1}^{(i)}(A)\underline{p}_0 \rangle_B &= \langle A\underline{p}_j, r_{m-1}^{(i)}(A)\underline{p}_0 \rangle_B \\ &= (\bar{\rho}_i)_{m-1} \langle A\underline{p}_j, \underline{p}_{m-1} \rangle_B + \dots + (\bar{\rho}_i)_0 \langle A\underline{p}_j, \underline{p}_0 \rangle_B \\ &= \langle \underline{\eta}_j, \underline{\rho}_i \rangle. \end{aligned}$$

Let $\{\underline{\phi}_0, \underline{\phi}_1, \dots, \underline{\phi}_{\kappa-1}\}$ be a basis for the range of $Q_B(A)$. It follows that

$$Q_B(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 = (\underline{\tau}_i)_0 \underline{\phi}_0 + \dots + (\underline{\tau}_i)_{\kappa-1} \underline{\phi}_{\kappa-1},$$

for some vector

$$\underline{\tau}_i = [(\underline{\tau}_i)_0, \dots, (\underline{\tau}_i)_{\kappa-1}]^T \in \mathcal{C}^\kappa. \quad (4.30)$$

Denote

$$\underline{\mu}_j = [\langle \underline{p}_j, \underline{\phi}_0 \rangle_B, \langle \underline{p}_j, \underline{\phi}_1 \rangle_B, \dots, \langle \underline{p}_j, \underline{\phi}_{\kappa-1} \rangle_B]^T \in \mathcal{C}^\kappa. \quad (4.31)$$

It follows that we can write

$$\langle \underline{p}_j, Q_B(A)\tilde{\psi}_{i-m}(A)\underline{p}_0 \rangle = \langle \underline{\mu}_j, \underline{\tau}_i \rangle.$$

Therefore, the entries of H_N simplify as:

$$h_{i,j} = \begin{cases} \frac{\langle \underline{\eta}_j, \underline{\rho}_i \rangle + \langle \underline{\mu}_j, \underline{\tau}_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j > \max\{i-1, i-m+\ell\} \\ \frac{\langle A\underline{p}_j, \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j = i, \dots, i-m+\ell \text{ (if } \ell \geq m) \\ 1 & j = i-1 \\ 0 & j < i-1 \end{cases}.$$

Let

$$\underline{w}_j = \begin{pmatrix} \eta_j \\ \underline{\mu}_j \end{pmatrix}, \quad \text{and} \quad \underline{v}_i = \frac{1}{\langle \underline{p}_i, \underline{p}_i \rangle_B} \begin{pmatrix} \rho_i \\ \underline{\tau}_i \end{pmatrix}, \quad (4.32)$$

be vectors of length $m + \kappa$, and notice that

$$\frac{\langle \eta_j, \rho_i \rangle + \langle \underline{\mu}_j, \underline{\tau}_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B} = \langle \underline{w}_j, \underline{v}_i \rangle.$$

Consider the following decomposition of the Hessenberg matrix H_N :

$$H_N = T + U,$$

where,

$$U = \begin{bmatrix} \langle \underline{w}_0, \underline{v}_0 \rangle & \cdots & \langle \underline{w}_{N-1}, \underline{v}_0 \rangle \\ & \ddots & \vdots \\ & & \langle \underline{w}_{N-1}, \underline{v}_{N-1} \rangle \end{bmatrix}, \quad (4.33)$$

and T is upper Hessenberg with entries,

$$t_{i,j} = \begin{cases} 0 & j > \max\{i-1, i-m+\ell\} \\ \frac{\langle A\underline{p}_j, \underline{p}_i \rangle_B - \langle \eta_j, \rho_i \rangle - \langle \underline{\mu}_j, \underline{\tau}_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j = i, \dots, i-m+\ell \quad (\text{if } \ell \geq m) \\ 1 & j = i-1 \\ 0 & j < i-1 \end{cases}.$$

Notice that T has an upper bandwidth of $\max\{0, \ell - m + 1\}$, so that if $m > \ell$, T has only a nonzero subdiagonal consisting of all ones. The matrix U can be further decomposed as

$$U = U_1 + U_2,$$

where,

$$U_1 = \begin{bmatrix} \frac{\langle \eta_0, \rho_0 \rangle}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \cdots & \frac{\langle \eta_{N-1}, \rho_0 \rangle}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ & \ddots & \vdots \\ & & \frac{\langle \eta_{N-1}, \rho_{N-1} \rangle}{\langle \underline{p}_{N-1}, \underline{p}_{N-1} \rangle_B} \end{bmatrix}, \quad U_2 = \begin{bmatrix} \frac{\langle \underline{\mu}_0, \underline{\tau}_0 \rangle}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \cdots & \frac{\langle \underline{\mu}_{N-1}, \underline{\tau}_0 \rangle}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ & \ddots & \vdots \\ & & \frac{\langle \underline{\mu}_{N-1}, \underline{\tau}_{N-1} \rangle}{\langle \underline{p}_{N-1}, \underline{p}_{N-1} \rangle_B} \end{bmatrix}.$$

Analogous to the B -normal(ℓ, m) case, the matrices U_1 and U_2 can be factored as follows:

$$\begin{aligned}
U_1 &= \begin{bmatrix} \frac{\underline{p}_0^T}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & & \\ & \ddots & \\ & & \frac{\underline{p}_{N-1}^T}{\langle \underline{p}_{N-1}, \underline{p}_{N-1} \rangle_B} \end{bmatrix} \begin{bmatrix} I_m & \cdots & I_m \\ & \ddots & \vdots \\ & & I_m \end{bmatrix} \begin{bmatrix} \underline{\eta}_0 & & \\ & \ddots & \\ & & \underline{\eta}_{N-1} \end{bmatrix} \\
&= M E D,
\end{aligned}$$

and

$$\begin{aligned}
U_2 &= \begin{bmatrix} \frac{\bar{p}_0^T}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & & \\ & \ddots & \\ & & \frac{\bar{p}_{N-1}^T}{\langle \underline{p}_{N-1}, \underline{p}_{N-1} \rangle_B} \end{bmatrix} \begin{bmatrix} I_\kappa & \cdots & I_\kappa \\ & \ddots & \vdots \\ & & I_\kappa \end{bmatrix} \begin{bmatrix} \underline{\mu}_0 & & \\ & \ddots & \\ & & \underline{\mu}_{N-1} \end{bmatrix} \\
&= \widehat{M} \widehat{E} \widehat{D}.
\end{aligned}$$

The dimensions of the matrices in the factorizations are as follows:

$$\begin{array}{ll}
M \text{ is } N \times (N \times m), & \widehat{M} \text{ is } N \times (N \times \kappa), \\
E \text{ is } (N \times m) \times (N \times m), & \widehat{E} \text{ is } (N \times \kappa) \times (N \times \kappa), \\
D \text{ is } (N \times m) \times N, & \widehat{D} \text{ is } (N \times \kappa) \times N.
\end{array}$$

The matrices E (\widehat{E}), consist of the blocks I_m (I_κ), on and above the diagonal. Each block I_m (I_κ) is an $m \times m$ ($\kappa \times \kappa$) identity matrix. Notice that E and \widehat{E} are both nonsingular. Except for the size, the structure of E^{-1} and \widehat{E}^{-1} , are the same as that given for the B -normal(ℓ, m) case in (4.19). They both have 1's on the main diagonal, but E^{-1} has -1 's on the $m + 1$ 'st superdiagonal, whereas \widehat{E}^{-1} has -1 's on the $\kappa + 1$ 'st superdiagonal.

Using the above factorizations for U_1 and U_2 , we obtain the matrix equation:

$$\begin{aligned}
AP_N &= P_N H_N \\
&= P_N T + P_N U_1 + P_N U_2 \\
&= P_N T + P_N M E D + P_N \widehat{M} \widehat{E} \widehat{D}.
\end{aligned}$$

Next, we define two matrices of auxiliary vectors:

$$\begin{aligned} Q &= P_N M E, \quad \text{and} \\ \widehat{Q} &= P_N \widehat{M} \widehat{E}. \end{aligned}$$

Making these substitutions into the above matrix equation, yields,

$$AP_N = P_N T + QD + \widehat{Q}\widehat{D}.$$

Notice that the dimensions of Q and \widehat{Q} are $N \times (N \times m)$ and $N \times (N \times \kappa)$, respectively.

In particular, we will denote

$$\begin{aligned} Q &= \left[\begin{array}{c|c|c|c|c|c|c|c|c|c} | & & | & | & & | & & | & & | \\ \hline \underline{q}_{0_0} & \cdots & \underline{q}_{0_{m-1}} & \underline{q}_{1_0} & \cdots & \underline{q}_{1_{m-1}} & \cdots & \underline{q}_{N-1_0} & \cdots & \underline{q}_{N-1_{m-1}} \\ \hline | & & | & | & & | & & | & & | \end{array} \right], \\ \widehat{Q} &= \left[\begin{array}{c|c|c|c|c|c|c|c|c|c} | & & | & | & & | & & | & & | \\ \hline \hat{\underline{q}}_{0_0} & \cdots & \hat{\underline{q}}_{0_{\kappa-1}} & \hat{\underline{q}}_{1_0} & \cdots & \hat{\underline{q}}_{1_{\kappa-1}} & \cdots & \hat{\underline{q}}_{N-1_0} & \cdots & \hat{\underline{q}}_{N-1_{\kappa-1}} \\ \hline | & & | & | & & | & & | & & | \end{array} \right]. \end{aligned}$$

Multiplying the matrix equations for Q and \widehat{Q} through by E^{-1} and \widehat{E}^{-1} , respectively, we obtain

$$\begin{aligned} QE^{-1} &= P_N M, \quad \text{and} \\ \widehat{Q}\widehat{E}^{-1} &= P_N \widehat{M}. \end{aligned}$$

It follows that the equations,

$$\begin{aligned} AP_N &= P_N T + QD + \widehat{Q}\widehat{D}, \\ QE^{-1} &= P_N M, \\ \widehat{Q}\widehat{E}^{-1} &= P_N \widehat{M}, \end{aligned} \tag{4.34}$$

define the matrix form of a set of multiple recursions that can be used to construct a B -orthogonal basis for $\mathcal{K}_N(\underline{p}_0, A)$. Equating the columns in the three matrix equations yields the formulas for the recursions at each step of the iteration. These

are given by:

$$\begin{aligned}
\underline{p}_{j+1} &= A\underline{p}_j - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{p}_i - \begin{bmatrix} \underline{q}_{j_0} & \cdots & \underline{q}_{j_{m-1}} \end{bmatrix} \underline{\eta}_j - \begin{bmatrix} \hat{\underline{q}}_{j_0} & \cdots & \hat{\underline{q}}_{j_{\kappa-1}} \end{bmatrix} \underline{\mu}_j, \\
\underline{q}_{j+1_i} &= \frac{(\bar{p}_{j+1})_i}{\langle \underline{p}_{j+1}, \underline{p}_{j+1} \rangle_B} \underline{p}_{j+1} + \underline{q}_{j_i}, \quad \text{for } i = 0, \dots, m-1, \\
\hat{\underline{q}}_{j+1_i} &= \frac{(\bar{\tau}_{j+1})_i}{\langle \underline{p}_{j+1}, \underline{p}_{j+1} \rangle_B} \underline{p}_{j+1} + \hat{\underline{q}}_{j_i}, \quad \text{for } i = 0, \dots, \kappa-1,
\end{aligned} \tag{4.35}$$

where, $\underline{\eta}_j$, $\underline{\rho}_j$, $\underline{\mu}_j$, and $\underline{\tau}_j$ are specified in (4.28-4.31), and the $t_{i,j}$'s are the elements of the Hessenberg matrix T . As in the B -normal(ℓ, m) case, these quantities stay bounded, thus, the problem of breakdown in the corresponding single (s, t)-recursion is avoided. More will be said about the actual computation of these quantities in the next chapter.

At each step, $m + \kappa + 1$ recursions are needed. The recursion for the direction vector \underline{p}_{j+1} involves one matrix vector multiplication with the previous direction vector, and $m + \kappa$ terms involving auxiliary vectors from the previous step. In addition, if $\ell \geq m$, it requires $\ell + m + 1$ previous direction vectors. The $m + \kappa$ recursions for the auxiliary vectors each involve only two terms. The work involved with this implementation is dependent on the degree of the polynomials p_ℓ , and q_m , as well as the rank of $Q_B(A)$.

This result is summarized in the following Theorem:

THEOREM 4.3 If there exists polynomials, $p_\ell(\lambda)$ and $q_m(\lambda)$, of degree ℓ and m respectively, such that

$$Q_B(A) = A^\dagger q_m(A) - p_\ell(A), \quad \text{and} \quad \text{Rank}(Q_B(A)) = \kappa, \tag{4.36}$$

then, for every \underline{x}_0 , a B -orthogonal basis for $\mathcal{K}_d(\underline{x}_0, A)$ can be constructed using the multiple recursions given in (4.35).

Proof: See above discussion. \square

Recall that if breakdown in the single (s, t) -recursion (4.10) occurs at step $j + 1$, the matrix given in (4.26) is singular. Notice from (4.29) and (4.31) that this matrix is equivalent to

$$\begin{bmatrix} \eta_{j-t} & \cdots & \eta_{j-1} \\ \mu_{j-t} & \cdots & \mu_{j-1} \end{bmatrix}. \quad (4.37)$$

What type of matrices satisfy (4.36)? Suppose the matrix A has the form:

$$A = \hat{A} + Z_r,$$

where \hat{A} is B -normal (ℓ, m) and Z_r is a rank r matrix, that is, A is a rank r perturbation of a B -normal (ℓ, m) matrix. It follows from Definition 4.1, that

$$\hat{A}^\dagger q_m(\hat{A}) - p_\ell(\hat{A}) = 0.$$

Notice that

$$q_m(A) = q_m(\hat{A} + Z_r) = a_m(\hat{A} + Z_r)^m + \cdots + a_1(\hat{A} + Z_r) + a_0 I.$$

By expanding each of these terms, we see that

$$q_m(A) = q_m(\hat{A}) + Z_{rm},$$

where Z_{rm} is an accumulation of all terms involving Z_r . It is easy to show that

$$\text{Range}(Z_{mr}) \subseteq \bigcup_{j=0}^{m-1} \text{Range}(\hat{A}^j Z_r).$$

Each of the terms involving Z_r has rank at most r . Since there are m of these terms, the rank of $Z_{rm} \leq rm$.

Similarly,

$$p_\ell(A) = p_\ell(\hat{A} + Z_r) = p_\ell(\hat{A}) + Z_{r\ell},$$

where the rank of $Z_{r\ell} \leq r\ell$. Using this information, we obtain

$$\begin{aligned} Q_B(A) &= A^\dagger q_m(A) - p_\ell(A) \\ &= (\hat{A}^\dagger + Z_r^\dagger)[q_m(\hat{A}) + Z_{rm}] - [p_\ell(\hat{A}) + Z_{r\ell}] \\ &= [\hat{A}^\dagger q_m(\hat{A}) - p_\ell(\hat{A})] + \hat{A}^\dagger Z_{rm} + Z_r^\dagger[q_m(\hat{A}) + Z_{rm}] - Z_{r\ell} \\ &= \hat{A}^\dagger Z_{rm} + Z_r^\dagger[q_m(\hat{A}) + Z_{rm}] - Z_{r\ell}, \end{aligned}$$

which yields

$$\text{Rank}(Q_B(A)) \leq (\ell + m + 1)r.$$

To summarize, if A is a rank r perturbation of a B -normal(ℓ, m) matrix, there exists polynomials, $p_\ell(\lambda)$ and $q_m(\lambda)$, of degree ℓ and m respectively, such that $A^\dagger q_m(A) - p_\ell(A) = Q_B(A)$, and,

$$\text{Rank}(Q_B(A)) \leq (\ell + m + 1)r.$$

COROLLARY 4.4 If $A = \hat{A} + Z_r$, where \hat{A} is B -normal(ℓ, m), and Z_r has rank r , then a B -orthogonal basis can be constructed using the multiple recursions given in (4.35), where $\kappa \leq (\ell + m + 1)r$.

Proof: The proof follows from Theorem 4.3 and the above discussion. \square

The recursions given by (4.35) could be applied to any low rank perturbation of a B -normal(ℓ, m) matrix. Since

$$\text{Rank}(Q_B(A)) = \kappa \leq (\ell + m + 1)r,$$

when r is large, κ can be large, and the number of recursions, and the number of terms needed to compute the basis increases. This means the practical application

of this form of recursion is limited to low rank perturbations of B -normal(ℓ, m) matrices.

From the previous section, we know that only certain B -normal(ℓ, m) matrices have more than a few distinct eigenvalues. Suppose $A = \hat{A} + Z_r$, where \hat{A} is B -normal(ℓ, m) with $d(\hat{A}) = k$, and Z_r has low rank r . There exists a polynomial p_k of degree k , such that $p_k(\hat{A}) = 0$. Notice that

$$p_k(\hat{A} + Z_r) = p_k(\hat{A}) + Z_{kr},$$

where the $\text{rank}(Z_{kr}) \leq kr$. It follows that there exists a polynomial q_{kr+1} of degree at most $(kr + 1)$, such that $q_{(kr+1)}(Z_{kr}) = 0$. By combining this information, we obtain

$$q_{(kr+1)}(p_k(\hat{A} + Z_r)) = 0,$$

which implies that $d(\hat{A} + Z_r) \leq k^2r + k$. This means that any low rank perturbation of a matrix with only a few distinct eigenvalues, still has only a few distinct eigenvalues. Therefore, only a few cases will be of interest. These are the matrices of the form

$$A = \hat{A} + Z_r,$$

where \hat{A} is either B -normal(1), B -unitary, or shifted B -unitary, and Z_r has low rank.

4.5 Breakdown In The Single (s, t) -Recursion

The next example illustrates that breakdown can occur during the (s, t) -iteration for B -normal(ℓ, m) matrices. Recall that breakdown occurs in the (s, t) -recursion for some \underline{p}_0 , if the system (4.13) is inconsistent at some step in the iteration.

Consider the unitary matrix

$$U = \begin{bmatrix} 0 & 0 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Since U is unitary with respect to the standard Euclidean inner product, $B = I$, and we have

$$U^\dagger = U^* = \frac{I}{U} = \frac{p_0(U)}{q_1(U)}.$$

By Definition 4.1, U is I -normal(0, 1). In the absence of breakdown, Section 4.3 shows an I -orthogonal basis $\{\underline{p}_j\}_{j=0}^4$ for $\mathcal{K}_5(\underline{p}_0, U)$ can be computed with the single (0, 1)-recursion,

$$\underline{p}_{j+1} = U\underline{p}_j + \beta_{j-1,j}U\underline{p}_{j-1} - \sigma_{j,j}\underline{p}_j, \quad \text{for } j = 0, \dots, 3.$$

We set $\beta_{-1,0} = 0$. If $j \geq 1$, $\beta_{j-1,j}$ is chosen to satisfy

$$\langle U\underline{p}_{j-1}, \underline{p}_0 \rangle \beta_{j-1,j} = -\langle U\underline{p}_j, \underline{p}_0 \rangle,$$

and then $\sigma_{j,j}$ is computed by:

$$\sigma_{j,j} = \frac{\langle U\underline{p}_j + \beta_{j-1,j}U\underline{p}_{j-1}, \underline{p}_j \rangle}{\langle \underline{p}_j, \underline{p}_j \rangle}.$$

Since an I -orthogonal basis is unique up to scale, the same basis can be obtained using any other recursion that constructs an I -orthogonal basis. For example, the full Gram-Schmidt process given by (2.8).

Let $\underline{p}_0 = [1, 0, 0, 0, 0]^T$. By computing $\{\underline{p}_j\}_{j=0}^4$ using (2.8), we obtain:

$$\begin{aligned} & \left\{ \begin{array}{c} | \\ \underline{p}_0 \\ | \end{array}, \begin{array}{c} | \\ \underline{p}_1 \\ | \end{array}, \begin{array}{c} | \\ \underline{p}_2 \\ | \end{array}, \begin{array}{c} | \\ \underline{p}_3 \\ | \end{array}, \begin{array}{c} | \\ \underline{p}_4 \\ | \end{array} \right\} \\ & = \left\{ \begin{array}{c} \left(\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} \right), \left(\begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{array} \right), \left(\begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{array} \right), \left(\begin{array}{c} 0 \\ 0 \\ 0 \\ 1/\sqrt{2} \\ 0 \end{array} \right), \left(\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 1/\sqrt{2} \end{array} \right) \end{array} \right\}. \end{aligned} \quad (4.38)$$

The computation of \underline{p}_1 is identical using the $(0, 1)$ -recursion. To compute \underline{p}_2 , we must choose $\beta_{0,1}$ to satisfy

$$\langle U_{\underline{p}_0}, \underline{p}_0 \rangle \beta_{0,1} = -\langle U_{\underline{p}_1}, \underline{p}_0 \rangle.$$

Since $\langle U_{\underline{p}_0}, \underline{p}_0 \rangle = \langle U_{\underline{p}_1}, \underline{p}_0 \rangle = 0$, $\beta_{0,1}$ can be chosen arbitrarily. Any choice of $\beta_{0,1}$ will yield $\underline{p}_2 = [0, 0, 1, 0, 0]^T$. Similarly, to compute \underline{p}_3 , we choose $\beta_{1,2}$ to satisfy

$$\langle U_{\underline{p}_1}, \underline{p}_0 \rangle \beta_{1,2} = -\langle U_{\underline{p}_2}, \underline{p}_0 \rangle,$$

Since $\langle U_{\underline{p}_1}, \underline{p}_0 \rangle = 0$, and $\langle U_{\underline{p}_2}, \underline{p}_0 \rangle = \frac{1}{\sqrt{2}}$, this is impossible, and breakdown occurs at this step. Recall from Section 4.3, that the multiple recursion given in (4.21) will also yield the basis vectors given by (4.38), without breakdown.

4.6 Concluding Remarks

This chapter considered general (s, t) -recursions of the form (4.10). In the absence of a condition called breakdown, a B -orthogonal basis can be constructed using this form of recursion when the matrix A is either B -normal(ℓ, m), or when there exists polynomials p_ℓ and q_m satisfying (4.36). Low rank perturbations of B -normal(ℓ, m) matrices are an example of the latter case.

Since breakdown is possible using the single (s, t) -recursion in exact arithmetic, and near breakdown can also pose numerical problems, a multiple recursion was developed that avoids the problem of breakdown. Sufficient conditions on the matrix A were given in Theorem 4.2 and Theorem 4.3 to guarantee that a B -orthogonal basis can be constructed using a multiple recursion for every \underline{p}_0 .

Breakdown corresponds to the inconsistency of the systems given in (4.13) and (4.26). In Chapter 6, we will show that these systems can be singular either for every \underline{p}_0 , or only for a set of initial vectors \underline{p}_0 of measure zero. If A is B -normal(ℓ, m), with $\ell, m \leq 1$, breakdown can be limited to a set of \underline{p}_0 of measure zero. This will

become important in establishing necessary conditions on the matrix A under which a multiple recursion will yield a B -orthogonal basis for every \underline{p}_0 .

5. Implementation And Numerical Results

5.1 Introduction

In Theorem 4.2 and Theorem 4.3, sufficient conditions on the matrix A were given in order that a B -orthogonal basis can be constructed using the multiple recursions given in (4.21) and (4.35), respectively. This chapter is concerned with the implementation of the conjugate gradient method using the above multiple recursions. Numerical examples are given that compare this implementation using short multiple recursions to that using a full conjugate gradient iteration, (Orthodir) (see Table 2.1).

5.2 B -Normal(ℓ, m) Matrices

We will first consider the construction of a B -orthogonal basis $\{\underline{p}_j\}_{j=0}^{d-1}$ for B -normal(ℓ, m) matrices using the multiple recursion given by:

$$\underline{p}_{j+1} = A\underline{p}_j - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{p}_i - \begin{bmatrix} \underline{q}_{j_0} & \underline{q}_{j_1} & \cdots & \underline{q}_{j_{m-1}} \end{bmatrix} \underline{\eta}_j, \quad (5.1)$$

$$\underline{q}_{j+1_i} = \frac{(\underline{p}_{j+1})_i}{\langle \underline{p}_{j+1}, \underline{p}_{j+1} \rangle_B} \underline{p}_{j+1} + \underline{q}_{j_i}, \quad \text{for } i = 0, \dots, m-1.$$

Recall from Section 4.3, that for any k ,

$$\underline{\eta}_k = [\langle A\underline{p}_k, \underline{p}_0 \rangle_B, \langle A\underline{p}_k, \underline{p}_1 \rangle_B, \dots, \langle A\underline{p}_k, \underline{p}_{m-1} \rangle_B]^T \in \mathcal{C}^m, \quad (5.2)$$

and,

$$\underline{\rho}_k = [(\underline{\rho}_k)_0, \dots, (\underline{\rho}_k)_{m-1}]^T \in \mathcal{C}^m, \quad (5.3)$$

is the vector of coefficients of the remainder term, $r_{m-1}^{(k)}(A)\underline{p}_0$, that results from dividing, $\underline{p}_k = \psi_k(A)\underline{p}_0$, by the polynomial

$$q_m(A) = b_m A^m + \cdots + b_1 A + b_0 I. \quad (5.4)$$

If $m \leq \ell$, the coefficients, $t_{i,j}$, are given by

$$t_{i,j} = \frac{\langle A\underline{p}_j, \underline{p}_i \rangle_B - \langle \underline{\eta}_j, \underline{\rho}_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B}, \quad \text{for } i = j - (\ell - m), \dots, j. \quad (5.5)$$

In this section, we will specify how these quantities can be computed.

The $\underline{\eta}_k$'s are obtained directly by computing the vector of inner products. If $\ell \geq m$, we must compute $t_{i,j}$, for $i = j - (\ell - m), \dots, j$. This could be done directly as specified by (5.5), but that would require storing $\underline{\rho}_i$, for $i = j - (\ell - m), \dots, j$, and the computation of additional inner products. Instead, at each step, we compute

$$\begin{aligned} \underline{y}_{j+1} &= A\underline{p}_j - [\underline{q}_{j_0} \cdots \underline{q}_{j_{m-1}}] \underline{\eta}_j, \\ \underline{p}_{j+1} &= \underline{y}_{j+1} - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{p}_i, \end{aligned} \quad (5.6)$$

$$\underline{q}_{j+1_i} = \frac{(\underline{\rho}_{j+1})_i}{\langle \underline{p}_{j+1}, \underline{p}_{j+1} \rangle_B} \underline{p}_{j+1} + \underline{q}_{j_i}, \quad \text{for } i = 0, \dots, m-1.$$

It is clear that the $t_{i,j}$'s must enforce B -orthogonality of \underline{p}_{j+1} to $\text{sp}\{\underline{p}_{j-(\ell-m)}, \dots, \underline{p}_j\}$.

This is accomplished more efficiently by computing,

$$t_{i,j} = \frac{\langle \underline{y}_{j+1}, \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B}, \quad \text{for } i = j - (\ell - m), \dots, j. \quad (5.7)$$

The $\underline{\rho}_k$'s, whose components $(\underline{\rho}_k)_i$ are used in the recursions for the auxiliary vectors, \underline{q}_{k_i} , for $i = 0, \dots, m-1$, are computed recursively. We will assume that A is B -normal(ℓ, m), with $m \neq 0$, since if $m = 0$, there is no remainder resulting upon division of $\underline{p}_k = \psi_k(A)\underline{p}_0$ by the constant polynomial $q_0(A)$. The vectors $\underline{\rho}_k = \underline{0}$, $\forall k$, and the multiple recursion simplifies to the single recursion,

$$\underline{p}_{j+1} = A\underline{p}_j - \sum_{i=j-\ell}^j t_{i,j} \underline{p}_i, \quad \text{where } t_{i,j} = \frac{\langle A\underline{p}_j, \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B},$$

for B -normal(ℓ) matrices [5].

Given an initial vector \underline{p}_0 , recall that \underline{p}_0 and each subsequent direction vector, \underline{p}_{j+1} , for $j = 0, 1, \dots$, can be written as

$$\underline{p}_{j+1} = \psi_{j+1}(A)\underline{p}_0,$$

for some polynomial ψ_{j+1} of exact degree $j + 1$. Each auxiliary vector, being the sum of the current direction vector and a previous auxiliary vector, can also be expressed as a polynomial in A times \underline{p}_0 . Given any vector \underline{z} that can be written as, $\underline{z} = \zeta(A)\underline{p}_0$, its remainder upon division by $q_m(A)$ is a polynomial in A , with degree at most $m - 1$, times \underline{p}_0 , which means it can be expressed as a linear combination of the \underline{p}_j 's, for $j = 0, \dots, m - 1$. We will denote this as,

$$\text{rem}(\underline{z}) = \alpha_0\underline{p}_0 + \dots + \alpha_{m-1}\underline{p}_{m-1},$$

and the corresponding vector of coefficients as,

$$\underline{rz} = [\alpha_0, \alpha_1, \dots, \alpha_{m-1}]^T \in \mathcal{C}^m.$$

If \underline{z} is expressed as a sum of polynomials in A times \underline{p}_0 ,

$$\underline{z} = \zeta(A)\underline{p}_0 = [\delta_1 \zeta_1(A) + \dots + \delta_s \zeta_s(A)]\underline{p}_0,$$

for some polynomials ζ_1, \dots, ζ_s , then the remainder could be obtained by dividing the individual polynomials by $q_m(A)$ and then adding the corresponding remainder terms. Notice that each of the recursions in (5.6) involves a sum of polynomials in A times \underline{p}_0 . This yields the following formulas for obtaining the remainder terms:

$$\begin{aligned} \text{rem}(\underline{y}_{j+1}) &= \text{rem}(A\underline{p}_j) - \sum_{i=1}^m (\underline{\eta}_j)_i \text{rem}(\underline{q}_{j_{i-1}}), \\ \text{rem}(\underline{p}_{j+1}) &= \text{rem}(\underline{y}_{j+1}) - \sum_{i=j-(\ell-m)}^j t_{i,j} \text{rem}(\underline{p}_i), \\ \text{rem}(\underline{q}_{j+1_i}) &= \frac{(\underline{p}_{j+1})_i}{\langle \underline{p}_{j+1}, \underline{p}_{j+1} \rangle_B} \text{rem}(\underline{p}_{j+1}) + \text{rem}(\underline{q}_{j_i}), \quad \text{for } i = 0, \dots, m - 1, \end{aligned}$$

and the recursions for the corresponding vectors of coefficients,

$$\begin{aligned}
\underline{ry}_{j+1} &= \underline{rAp}_j - \sum_{i=1}^m (\underline{\eta}_j)_i \underline{rq}_{j_{i-1}}, \\
\underline{rp}_{j+1} &= \underline{ry}_{j+1} - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{rp}_i, \\
\underline{rq}_{j+1_i} &= \frac{(\underline{\rho}_{j+1})_i}{\langle \underline{p}_{j+1}, \underline{p}_{j+1} \rangle_B} \underline{rp}_{j+1} + \underline{rq}_{j_i}, \quad \text{for } i = 0, \dots, m-1.
\end{aligned} \tag{5.8}$$

We note here that

$$\underline{\rho}_k = \underline{rp}_k, \quad \forall k.$$

Some startup information is required to begin the recursive computation for the $\underline{\rho}_k$'s. This information can be obtained by computing the first $m+1$ direction vectors $\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_m\}$ using a full Gram-Schmidt process (2.8), or some modified version of it. This yields the corresponding matrix equation,

$$AP_m = P_{m+1}H_{m+1,m}. \tag{5.9}$$

The Hessenberg matrix $H_{m+1,m}$ is formed and stored for later use in the recursive update of the $\underline{\rho}_k$'s. In addition to computing the direction vectors, the \underline{q}_{k_i} 's and the remainder terms (5.8) will also need to be updated during these steps for later use in the iteration.

No computation is necessary to obtain, $\underline{\rho}_0, \dots, \underline{\rho}_{m-1}$. First, $\underline{p}_0 = \underline{r}_0 = \psi_0(A)\underline{p}_0$. Notice that the remainder of \underline{p}_0 upon division by $q_m(A)$ is itself, $\text{rem}(\underline{p}_0) = 1 \underline{p}_0$, which yields,

$$\underline{rp}_0 = \underline{\rho}_0 = [1 \ 0 \ \dots \ 0]^T \in \mathcal{C}^m.$$

The auxiliary vectors and the corresponding remainder terms are computed as:

$$\underline{q}_{0_i} = \frac{(\underline{\rho}_0)_i}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \underline{p}_0, \quad \underline{rq}_{0_i} = \frac{(\underline{\rho}_0)_i}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \underline{\rho}_0, \quad \text{for } i = 0, \dots, m-1.$$

Similarly, for $j = 0, \dots, m-2$ $\text{rem}(\underline{p}_{j+1}) = \underline{p}_{j+1}$, which yields,

$$\underline{rp}_{j+1} = \underline{\rho}_{j+1} = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T \in \mathcal{C}^m,$$

which is a vector of all zeros, except for a 1 in the $j + 1$ 'st position. The auxiliary vectors and their corresponding remainder terms can be computed using (5.6) and (5.8).

The computation of $\underline{\rho}_m$ is more complicated. To facilitate this, we write the polynomial $q_m(A)$ given in (5.4) as a linear combination of the ψ_k 's, for $k = 0, \dots, m$,

$$\begin{aligned} q_m(A) &= \gamma_m \psi_m(A) + \dots + \gamma_1 \psi_1(A) + \gamma_0 \psi_0(A) \\ &= b_m A^m + \dots + b_0 I. \end{aligned} \tag{5.10}$$

This requires that we accumulate the polynomials, ψ_0, \dots, ψ_m , during the iteration, where,

$$\underline{p}_{j+1} = \psi_{j+1}(A)\underline{p}_0 = [\alpha_{j+1}A^{j+1} + \alpha_j A^j + \dots + \alpha_1 A + \alpha_0 I]\underline{p}_0.$$

The coefficients of $\psi_{j+1}(\mu)$ are stored in the vector of length $m + 1$,

$$\underline{cp}_{j+1} = [\alpha_0, \alpha_1, \dots, \alpha_j, \alpha_{j+1}, 0, \dots, 0]^T.$$

Since

$$A\psi_{j+1}(A)\underline{p}_0 = [\alpha_{j+1}A^{j+2} + \alpha_j A^{j+1} + \dots + \alpha_1 A^2 + \alpha_0 A]\underline{p}_0,$$

the corresponding vector of coefficients is obtained by shifting the vector \underline{cp}_{j+1} to the right one place, i.e.,

$$\underline{cAp}_{j+1} = [0, \alpha_0, \alpha_1, \dots, \alpha_j, \alpha_{j+1}, 0, \dots, 0]_{m+1}^T.$$

Starting with $\underline{p}_0 = \psi_0(A)\underline{p}_0 = 1\underline{p}_0$, we obtain

$$\underline{cp}_0 = [1 \ 0 \ \dots \ 0]_{m+1}^T.$$

Using (2.8), we see that a recursion for the vector of coefficients for the polynomials, ψ_{j+1} , for $j = 0, \dots, m - 1$, is given by,

$$\underline{cp}_{j+1} = \underline{cAp}_j - \sum_{i=0}^j \sigma_{i,j} \underline{cp}_i.$$

The vectors, $\underline{cp}_0, \dots, \underline{cp}_m$, are stored in the matrix K , which has the following form:

$$K = \left[\begin{array}{c|c|c|c|c} | & | & & | & \\ \hline \underline{cp}_0 & \underline{cp}_1 & \cdots & \underline{cp}_m & \\ \hline | & | & & | & \end{array} \right] = \begin{bmatrix} 1 & * & \cdots & \cdots & * \\ 0 & 1 & \ddots & & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & * \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix},$$

where the *'s denote possible nonzero entries. Finally, the γ 's in (5.10) can be obtained by solving the system,

$$K \begin{pmatrix} \gamma_0 \\ \vdots \\ \gamma_m \end{pmatrix} = \begin{pmatrix} b_0 \\ \vdots \\ b_m \end{pmatrix},$$

using backsubstitution.

Using this form for $q_m(A)$, it is easy to see that the remainder of \underline{p}_m upon division by $q_m(A)$ is given by,

$$\text{rem}(\underline{p}_m) = -\frac{\gamma_{m-1}}{\gamma_m} \underline{p}_{m-1} - \cdots - \frac{\gamma_1}{\gamma_m} \underline{p}_1 - \frac{\gamma_0}{\gamma_m} \underline{p}_0,$$

and thus,

$$\underline{\rho}_m = \left[-\frac{\gamma_0}{\gamma_m}, -\frac{\gamma_1}{\gamma_m}, \dots, -\frac{\gamma_{m-1}}{\gamma_m} \right]^T. \quad (5.11)$$

Once $\underline{\rho}_m$ is computed, \underline{q}_{m_i} , and \underline{rq}_{m_i} , for $i = 0, \dots, m-1$, can be computed according to (5.6) and (5.8).

When $j = m$, we can begin the recursive update of $\underline{rp}_{j+1} = \underline{\rho}_{j+1}$ using (5.8). First, notice that

$$\begin{aligned} \text{rem}(A\underline{p}_j) &= \text{rem}(A\psi_j(A)\underline{p}_0) \\ &= \text{rem} \left(A \left[\tilde{\psi}_{j-m}(A)q_m(A)\underline{p}_0 + r_{m-1}^{(j)}(A)\underline{p}_0 \right] \right) \\ &= \text{rem} (A \text{rem}(\underline{p}_j)). \end{aligned}$$

Since $\text{rem}(\underline{p}_j) = (\underline{\rho}_j)_0 \underline{p}_0 + \cdots + (\underline{\rho}_j)_{m-1} \underline{p}_{m-1}$, the matrix equation in (5.9) can be used to write,

$$\begin{aligned} A \text{rem}(\underline{p}_j) &= A[\underline{p}_0, \dots, \underline{p}_{m-1}] \underline{\rho}_j = [\underline{p}_0 \cdots \underline{p}_m] H_{m+1,m} \underline{\rho}_j \\ &= c_0 \underline{p}_0 + c_1 \underline{p}_1 + \cdots + c_m \underline{p}_m, \end{aligned}$$

where,

$$\underline{c}_j = [c_0, c_1, \dots, c_m]^T = H_{m+1,m} \underline{\rho}_j.$$

Therefore,

$$\begin{aligned} \text{rem}(A \underline{p}_j) &= \text{rem}(A \text{rem}(\underline{p}_j)) \\ &= c_0 \text{rem}(\underline{p}_0) + c_1 \text{rem}(\underline{p}_1) + \cdots + c_m \text{rem}(\underline{p}_m) \\ &= c_0 \underline{p}_0 + c_1 \underline{p}_1 + \cdots + c_m \left[-\frac{\gamma_{m-1}}{\gamma_m} \underline{p}_{m-1} - \cdots - \frac{\gamma_0}{\gamma_m} \underline{p}_0 \right], \end{aligned}$$

which yields

$$\underline{r} A \underline{p}_j = c_0 \underline{\rho}_0 + c_1 \underline{\rho}_1 + \cdots + c_m \underline{\rho}_m.$$

When $j \geq m$, we can compute \underline{p}_{j+1} and the auxiliary vectors using (5.6), and the remainder terms can be updated recursively using (5.8).

This process describes the construction of a B -orthogonal basis for B -normal(ℓ, m) matrices. To obtain a conjugate gradient algorithm for $\mathcal{CG}(B, A)$, at each step, we compute the additional recursions:

$$\begin{aligned} \underline{x}_{j+1} &= \underline{x}_j + \alpha_j \underline{p}_j, & \alpha_j &= \frac{\langle B \underline{e}_j, \underline{p}_j \rangle}{\langle B \underline{p}_j, \underline{p}_j \rangle}, \\ \underline{r}_{j+1} &= \underline{r}_j - \alpha_j A \underline{p}_j. \end{aligned} \tag{5.12}$$

For clarity of the order of computations, we outline the algorithm below.

ALGORITHM 5.1 CG algorithm for B -normal(ℓ, m) matrices:

Input: A , \underline{x}_0 , $\underline{r}_0 = b - A \underline{x}_0$, coefficients of $q_m(A) = \underline{\beta} = [b_m, \dots, b_0]^T$.

$$\begin{aligned} \underline{p}_0 &= \underline{r}_0, \\ \underline{c} \underline{p}_0 &= [1 \ 0 \ \cdots \ 0]_{m+1}^T, & \underline{\rho}_0 &= [1 \ 0 \ \cdots \ 0]_m^T, \end{aligned}$$

$$\underline{q}_{0_i} = \frac{(\underline{\rho}_0)_i}{\langle B\underline{p}_0, \underline{p}_0 \rangle} \underline{p}_0, \quad \underline{r}q_{0_i} = \frac{(\underline{\rho}_0)_i}{\langle B\underline{p}_0, \underline{p}_0 \rangle} \underline{\rho}_0, \quad i = 0, \dots, m-1,$$

for $j = 0, \dots, m-1$

$$\underline{x}_{j+1} = \underline{x}_j + \alpha_j \underline{p}_j, \quad \underline{r}_{j+1} = \underline{r}_j - \alpha_j A \underline{p}_j, \quad \alpha_j = \frac{\langle B \underline{e}_j, \underline{p}_j \rangle}{\langle B \underline{p}_j, \underline{p}_j \rangle},$$

$$\underline{p}_{j+1} = A \underline{p}_j - \sum_{i=0}^j \sigma_{i,j} \underline{p}_i, \quad \sigma_{i,j} = \frac{\langle B A \underline{p}_j, \underline{p}_i \rangle}{\langle B \underline{p}_i, \underline{p}_i \rangle},$$

$$\underline{c}p_{j+1} = \underline{c}A \underline{p}_j - \sum_{i=0}^j \sigma_{i,j} \underline{c}p_i,$$

if $j < m-1$

$$\underline{\rho}_{j+1} = [0 \cdots 0 \ 1 \ 0 \cdots 0]_m^T,$$

if $j = m-1$

$$\underline{\gamma} = K \setminus \underline{\beta},$$

$$\underline{\rho}_{j+1} = \left[-\frac{\gamma_0}{\gamma_m}, \dots, -\frac{\gamma_{m-1}}{\gamma_m} \right]_m^T,$$

$$\underline{q}_{j+1_i} = \frac{(\underline{\rho}_{j+1})_i}{\langle B \underline{p}_{j+1}, \underline{p}_{j+1} \rangle} \underline{p}_{j+1} + \underline{q}_{j_i}, \quad i = 0, \dots, m-1,$$

$$\underline{r}q_{j+1_i} = \frac{(\underline{\rho}_{j+1})_i}{\langle B \underline{p}_{j+1}, \underline{p}_{j+1} \rangle} \underline{\rho}_{j+1} + \underline{r}q_{j_i}, \quad i = 0, \dots, m-1,$$

for $j = m, \dots$

$$\underline{x}_{j+1} = \underline{x}_j + \alpha_j \underline{p}_j, \quad \underline{r}_{j+1} = \underline{r}_j - \alpha_j A \underline{p}_j, \quad \alpha_j = \frac{\langle B \underline{e}_j, \underline{p}_j \rangle}{\langle B \underline{p}_j, \underline{p}_j \rangle},$$

$$\underline{y}_{j+1} = A \underline{p}_j - \sum_{i=1}^m (\underline{\eta}_j)_i \underline{q}_{j_{i-1}}, \quad \underline{\eta}_j = [\langle A \underline{p}_j, \underline{p}_0 \rangle_B, \dots, \langle A \underline{p}_j, \underline{p}_{m-1} \rangle_B]_m^T,$$

$$\underline{p}_{j+1} = \underline{y}_{j+1} - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{p}_i, \quad t_{i,j} = \frac{\langle B \underline{y}_{j+1}, \underline{p}_i \rangle}{\langle B \underline{p}_i, \underline{p}_i \rangle},$$

$$\underline{c}_j = H_{m+1,m} \underline{\rho}_j, \quad \underline{r}A \underline{p}_j = c_0 \underline{\rho}_0 + \cdots + c_m \underline{\rho}_m,$$

$$\underline{r}y_{j+1} = \underline{r}A \underline{p}_j - \sum_{i=1}^m (\underline{\eta}_j)_i \underline{r}q_{j_{i-1}}, \quad \underline{\rho}_{j+1} = \underline{r}y_{j+1} - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{\rho}_i,$$

$$\underline{q}_{j+1_i} = \frac{(\underline{\rho}_{j+1})_i}{\langle B \underline{p}_{j+1}, \underline{p}_{j+1} \rangle} \underline{p}_{j+1} + \underline{q}_{j_i}, \quad i = 0, \dots, m-1,$$

$$\underline{r}q_{j+1_i} = \frac{(\underline{\rho}_{j+1})_i}{\langle B \underline{p}_{j+1}, \underline{p}_{j+1} \rangle} \underline{\rho}_{j+1} + \underline{r}q_{j_i}, \quad i = 0, \dots, m-1.$$

Recall from Section 4.2 that if A is B -normal(ℓ, m), and either ℓ or m is greater than 1, then A has a relatively small number ($\max\{\ell^2, m^2\} + 1$) of distinct eigenvalues. For this reason, we are mainly interested in the cases when $(\ell, m) \leq (1, 1)$. Although there are already economical conjugate gradient algorithms for these cases, we will illustrate the algorithm with an example.

First, if A is I -normal(ℓ, m), $A^*A = AA^*$, and

$$A^*q_m(A) - p_\ell(A) = 0,$$

for some polynomials p_ℓ and q_m of degrees ℓ and m , respectively. Notice that if $B = A^*A$,

$$A^\dagger = B^{-1}A^*B = A^*,$$

and it follows that

$$A^\dagger q_m(A) - p_\ell(A) = 0,$$

and A is B -normal(ℓ, m), with $B = A^*A$.

Example 1: Consider the shifted unitary matrix given by,

$$A = 10 U + (8 + 8i) I.$$

This matrix is I -normal(1, 1). From the above, A is also B -normal(1, 1), for $B = A^*A$. Thus, a minimal residual method, $\mathcal{CG}(A^*A, A)$, can be implemented using Algorithm 5.1. The eigenvalues of A are plotted in Figure 5.1. In Figure 5.2, we compare the convergence measured in the relative residual norm of Algorithm 5.1 to that obtained by using a full conjugate gradient iteration, Orthodir, given in Table 2.1. The curves lay exactly on top of each other, as our theory predicts.

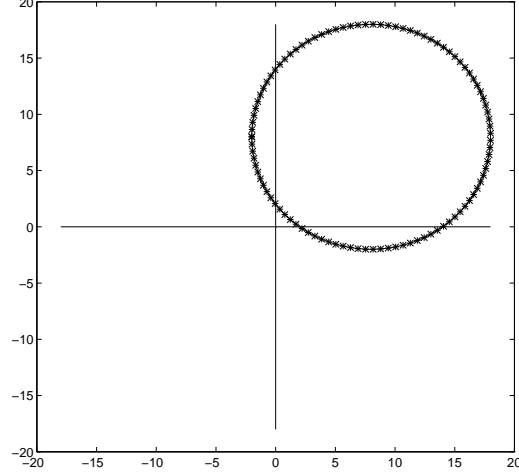


Figure 5.1: Spectrum of problem given in Example 1.

5.3 Generalizations Of B -Normal(ℓ, m) Matrices

In Section 4.4 it was demonstrated that if polynomials, p_ℓ and q_m , of degrees ℓ and m respectively exist so that

$$Q_B(A) = A^\dagger q_m(A) - p_\ell(A),$$

and,

$$\text{Rank}(Q_B(A)) = \kappa,$$

then a B -orthogonal basis could be constructed using the multiple recursion:

$$\begin{aligned} \underline{p}_{j+1} &= A\underline{p}_j - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{p}_i - [\underline{q}_{j_0} \cdots \underline{q}_{j_{m-1}}] \underline{\eta}_j - [\hat{\underline{q}}_{j_0} \cdots \hat{\underline{q}}_{j_{\kappa-1}}] \underline{\mu}_j, \\ \underline{q}_{j+1_i} &= \frac{(\bar{\underline{p}}_{j+1})^i}{\langle \underline{p}_{j+1}, \underline{p}_{j+1} \rangle_B} \underline{p}_{j+1} + \underline{q}_{j_i}, \quad \text{for } i = 0, \dots, m-1, \\ \hat{\underline{q}}_{j+1_i} &= \frac{(\bar{\hat{\underline{p}}}_{j+1})^i}{\langle \hat{\underline{p}}_{j+1}, \hat{\underline{p}}_{j+1} \rangle_B} \hat{\underline{p}}_{j+1} + \hat{\underline{q}}_{j_i}, \quad \text{for } i = 0, \dots, \kappa-1. \end{aligned} \tag{5.13}$$

For any k ,

$$\begin{aligned} \underline{\eta}_k &= [\langle A\underline{p}_k, \underline{p}_0 \rangle_B, \dots, \langle A\underline{p}_k, \underline{p}_{m-1} \rangle_B]^T \in \mathcal{C}^m, \\ \underline{\mu}_k &= [\langle \underline{p}_k, \hat{\underline{p}}_0 \rangle_B, \dots, \langle \underline{p}_k, \hat{\underline{p}}_{\kappa-1} \rangle_B]^T \in \mathcal{C}^\kappa, \end{aligned} \tag{5.14}$$

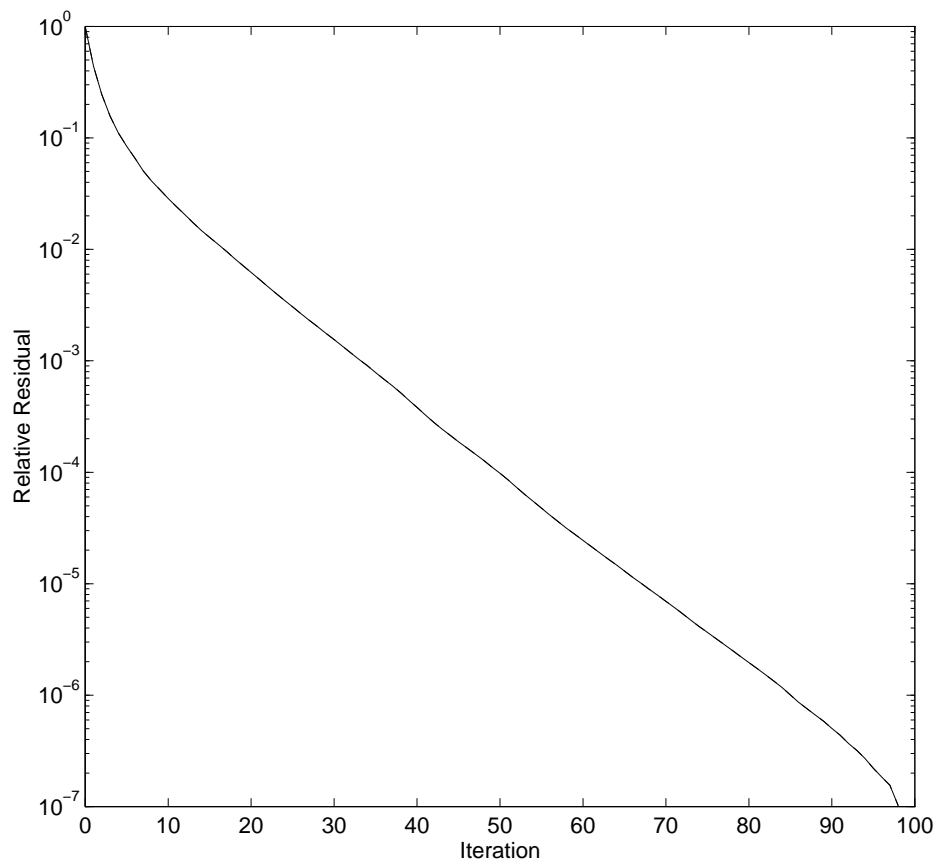


Figure 5.2. Convergence of Orthodir (solid line), Algorithm 5.1 (dashed line, plotted on top of solid line), on Example 1.

where the vectors $\{\underline{\phi}_0, \dots, \underline{\phi}_{\kappa-1}\}$ form a basis for the range of $Q_B(A)$. If $m \leq \ell$, the coefficients $t_{i,j}$ are given by,

$$t_{i,j} = \frac{\langle A\underline{p}_j, \underline{p}_i \rangle_B - \langle \underline{\eta}_j, \underline{\rho}_i \rangle - \langle \underline{\mu}_j, \underline{\tau}_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B}, \quad \text{for } i = j - (\ell - m), \dots, j. \quad (5.15)$$

The vector

$$\underline{\rho}_k = [(\underline{\rho}_k)_0, \dots, (\underline{\rho}_k)_{m-1}]^T \in \mathcal{C}^m,$$

is the vector of coefficients of the remainder term, $r_{m-1}^{(k)}(A)\underline{p}_0$, that results from dividing \underline{p}_k by $q_m(A)$, and $Q_B(A)\tilde{\psi}_{k-m}(A)\underline{p}_0 = (\underline{\tau}_k)_0\underline{\phi}_0 + \dots + (\underline{\tau}_k)_{\kappa-1}\underline{\phi}_{\kappa-1}$, for some vector of coefficients,

$$\underline{\tau}_k = [(\underline{\tau}_k)_0, \dots, (\underline{\tau}_k)_{\kappa-1}]^T \in \mathcal{C}^\kappa.$$

In this section, we will be concerned with computing a basis for the range of $Q_B(A)$, as well as computing the $t_{i,j}$'s, and the quantities: $\underline{\eta}_k$, $\underline{\mu}_k$, $\underline{\rho}_k$, and $\underline{\tau}_k$. The order of the computations will become important as the information needed to compute them may not be available until some later step in the iteration.

Recall that $A^\dagger = B^{-1}A^*B$, so that

$$Q_B(A) = (B^{-1}A^*B) q_m(A) - p_\ell(A).$$

A basis for the range of $Q_B(A)$ could be obtained by first computing $Q_B(A)$, then, since $Q_B(A)$ is rank deficient, a basis for its range can be obtained using a QR decomposition routine with column pivoting (see [9], pp. 233-236). The computation of $Q_B(A)$ and a basis for its range may be very simple in some cases. For example, if

$$A = \hat{A} + Z_r,$$

where \hat{A} is I -self-adjoint, and Z_r is a rank r matrix that can be written as an outer product,

$$Z_r = W_r V_r^*,$$

for some $N \times r$ matrices V_r and W_r of rank r . Since \widehat{A} is I -self-adjoint,

$$\widehat{A}^* = \frac{p_\ell(\widehat{A})}{q_m(\widehat{A})} = \widehat{A}.$$

This yields $q_m(\widehat{A}) = I$, and $p_\ell(\widehat{A}) = \widehat{A}$. Let $B = A^*A$. Notice that

$$A^\dagger = B^{-1}A^*B = (A^*A)^{-1}A^*A^*A = A^{-1}A^*A,$$

and

$$\begin{aligned} Q_{A^*A}(A) &= A^\dagger q_m(A) - p_\ell(A) = A^\dagger - A \\ &= A^{-1}A^*A - A = A^{-1}(A^* - A)A \\ &= A^{-1}(V_r W_r^* - W_r V_r^*)A. \end{aligned}$$

Denote $\{\widehat{v}_i\}_{i=1}^n$ as the linearly independent columns from the matrices V_r and W_r .

Now,

$$\text{Range}(Q_{A^*A}(A)) = \text{sp}\{A^{-1}\widehat{v}_i\}_{i=1}^n = \text{sp}\{\underline{\phi}_0, \dots, \underline{\phi}_{\kappa-1}\}. \quad (5.16)$$

The algorithm will only require that we compute the vectors

$$\begin{aligned} \underline{\mu}_k &= \left[\langle \underline{p}_k, \underline{\phi}_0 \rangle_B, \dots, \langle \underline{p}_k, \underline{\phi}_{\kappa-1} \rangle_B \right]^T, \quad \text{where} \\ \langle \underline{p}_k, \underline{\phi}_i \rangle_B &= \langle A^*A \underline{p}_k, \underline{\phi}_i \rangle = \langle A \underline{p}_k, A(A^{-1}\widehat{v}_i) \rangle = \langle A \underline{p}_k, \widehat{v}_i \rangle. \end{aligned} \quad (5.17)$$

The quantities $\underline{\eta}_k$ and $\underline{\mu}_k$ are computed as specified by (5.14). Analogous to the B -normal(ℓ, m) case, the $\underline{\rho}_k$'s can be computed recursively after the information needed to start the recursion is available.

At any step $k + 1$, the recursion for the direction vector in (5.13) involves both sets of auxiliary vectors, \underline{q}_{k_i} 's and $\underline{\hat{q}}_{k_i}$'s. Before we can compute the $\underline{\hat{q}}_{k_i}$'s, we must know the vector $\underline{\tau}_k$. The method we will describe for computing this vector is not possible at step $k + 1$, and will actually be done at a later step in the iteration. This means that the multiple recursion (5.13) must be modified in a way that allows the computation of the auxiliary vectors to be delayed. After describing how to compute $\underline{\tau}_k$, we will show how this can be done. An outline of the algorithm will then be given to clarify the order of the computations.

Recall from the development in Section 4.4 that led to the multiple recursions in (5.13), that the entries in the Hessenberg matrix H_N simplify yielding,

$$h_{k,j} = \frac{\langle A\underline{p}_j, \underline{p}_k \rangle_B}{\langle \underline{p}_k, \underline{p}_k \rangle_B} = \frac{\langle \underline{\eta}_j, \underline{\rho}_k \rangle + \langle \underline{\mu}_j, \underline{\tau}_k \rangle}{\langle \underline{p}_k, \underline{p}_k \rangle_B}, \quad \text{if } j > \max\{k-1, k-m+\ell\}. \quad (5.18)$$

Denote

$$\vartheta = \begin{cases} 1 & \text{if } m > \ell \\ m - \ell & \text{if } m \leq \ell \end{cases}.$$

Notice that if $j > k - \vartheta$, we can compute,

$$\langle \underline{\mu}_j, \underline{\tau}_k \rangle = \langle A\underline{p}_j, \underline{p}_k \rangle_B - \langle \underline{\eta}_j, \underline{\rho}_k \rangle.$$

All the entries to the right of column $(k - \vartheta)$ in the k 'th row of H_N involve the same vector $\underline{\tau}_k$. By moving to the right κ places, we can obtain κ equations involving $\underline{\tau}_k$.

At step $j + 1 = k - \vartheta + \kappa + 1$, we obtain the system,

$$\begin{bmatrix} - & \underline{\mu}_j & - \\ & \vdots & \\ - & \underline{\mu}_{j-\kappa+1} & - \end{bmatrix} \bar{\underline{\tau}}_k = \begin{pmatrix} \langle A\underline{p}_j, \underline{p}_k \rangle_B - \langle \underline{\eta}_j, \underline{\rho}_k \rangle \\ \vdots \\ \langle A\underline{p}_{j-\kappa+1}, \underline{p}_k \rangle_B - \langle \underline{\eta}_{j-\kappa+1}, \underline{\rho}_k \rangle \end{pmatrix} \quad (5.19)$$

for $\bar{\underline{\tau}}_k$. For generalizations of B -normal (ℓ, m) matrices, theory guarantees that (5.18) holds. Thus, if we have chosen $\{\underline{\phi}_0, \dots, \underline{\phi}_{\kappa-1}\}$ properly, that is, if they form a basis for the range of $Q_B(A)$, it follows that the above system is consistent.

From the above, we see that when $j = \kappa - \vartheta$, we can compute $\bar{\underline{\tau}}_0$ using (5.19), and when $j = \kappa - \vartheta + 1$, we can compute $\bar{\underline{\tau}}_1$, and so forth. Denote

$$\theta = \kappa - \vartheta,$$

and notice that whenever $j \geq \theta$, $\bar{\underline{\tau}}_{j-\theta}$ is computable using (5.19). The auxiliary vectors can then be computed as:

$$\underline{q}_{j-\theta_i} = \frac{(\bar{\underline{p}}_{j-\theta})_i}{\langle \underline{p}_{j-\theta}, \underline{p}_{j-\theta} \rangle_B} \underline{p}_{j-\theta} + \underline{q}_{j-\theta_i}, \quad \text{for } i = 0, \dots, m-1, \quad (5.20)$$

$$\hat{\underline{q}}_{j-\theta_i} = \frac{(\bar{\underline{\tau}}_{j-\theta})_i}{\langle \underline{p}_{j-\theta}, \underline{p}_{j-\theta} \rangle_B} \underline{p}_{j-\theta} + \hat{\underline{q}}_{j-\theta_i}, \quad \text{for } i = 0, \dots, \kappa-1.$$

However, at this step, the multiple recursion in (5.13) requires us to compute

$$\underline{p}_{j+1} = A\underline{p}_j - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{p}_i - [\underline{q}_{j_0} \cdots \underline{q}_{j_{m-1}}] \underline{\eta}_j - [\hat{\underline{q}}_{j_0} \cdots \hat{\underline{q}}_{j_{\kappa-1}}] \underline{\mu}_j.$$

Unless $\theta = 0$, $\bar{\underline{x}}_j$ can't be computed using (5.19). Substituting in the quantities,

$$\underline{q}_{j_i} = \frac{(\bar{\rho}_j)_i}{\langle \underline{p}_j, \underline{p}_j \rangle_B} \underline{p}_j + \underline{q}_{j-1_i}, \quad \text{for } i = 0, \dots, m-1, \quad \text{and}$$

$$\hat{\underline{q}}_{j_i} = \frac{(\bar{\tau}_j)_i}{\langle \underline{p}_j, \underline{p}_j \rangle_B} \underline{p}_j + \hat{\underline{q}}_{j-1_i}, \quad \text{for } i = 0, \dots, \kappa-1,$$

from the previous step yields,

$$\underline{p}_{j+1} = A\underline{p}_j - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{p}_i - \delta_j \underline{p}_j - [\underline{q}_{j-1_0} \cdots \underline{q}_{j-1_{m-1}}] \underline{\eta}_j - [\hat{\underline{q}}_{j-1_0} \cdots \hat{\underline{q}}_{j-1_{\kappa-1}}] \underline{\mu}_j,$$

for some constant δ_j which will be determined later. This process of substituting in auxiliary vectors from previous steps can be repeated until we obtain:

$$\begin{aligned} \underline{p}_{j+1} &= A\underline{p}_j - \sum_{i=j-(\ell-m)}^j t_{i,j} \underline{p}_i - \delta_j \underline{p}_j - \cdots - \delta_{j-\theta+1} \underline{p}_{j-\theta+1} \\ &\quad - [\underline{q}_{j-\theta_0} \cdots \underline{q}_{j-\theta_{m-1}}] \underline{\eta}_j - [\hat{\underline{q}}_{j-\theta_0} \cdots \hat{\underline{q}}_{j-\theta_{\kappa-1}}] \underline{\mu}_j, \end{aligned} \quad (5.21)$$

where, $\bar{\underline{x}}_{j-\theta}$ can be computed using (5.19), and the, $\underline{q}_{j-\theta_i}$'s and $\hat{\underline{q}}_{j-\theta_i}$'s, can be computed using (5.20). The $t_{i,j}$'s are given by (5.15), and

$$\delta_n = \left(\sum_{i=0}^{m-1} \frac{(\bar{\rho}_n)_i}{\langle \underline{p}_n, \underline{p}_n \rangle_B} (\underline{\eta}_j)_i + \sum_{i=0}^{\kappa-1} \frac{(\bar{\tau}_n)_i}{\langle \underline{p}_n, \underline{p}_n \rangle_B} (\underline{\mu}_j)_i \right), \quad \text{for } n = j - \theta + 1, \dots, j.$$

At first glance, it looks like the information needed to compute the α_n 's and the $t_{i,j}$'s is not available. Let

$$\varphi = \begin{cases} j - \theta + 1 & \text{if } m > \ell \\ \min\{j - (\ell - m), j - \theta + 1\} & \text{if } m \leq \ell \end{cases}. \quad (5.22)$$

Instead of calculating the α_n 's and $t_{i,j}$'s as specified, we rewrite (5.21) as,

$$\underline{y}_{j+1} = A\underline{p}_j - [\underline{q}_{j-\theta_0} \cdots \underline{q}_{j-\theta_{m-1}}] \underline{\eta}_j - [\hat{\underline{q}}_{j-\theta_0} \cdots \hat{\underline{q}}_{j-\theta_{\kappa-1}}] \underline{\mu}_j, \quad (5.23)$$

$$\underline{p}_{j+1} = \underline{y}_{j+1} - \sum_{i=\varphi}^j \hat{t}_{i,j} \underline{p}_i,$$

where the $\hat{t}_{i,j}$'s must enforce B -orthogonality of \underline{p}_{j+1} to $\text{sp}\{\underline{p}_\varphi, \dots, \underline{p}_j\}$, which yields,

$$\hat{t}_{i,j} = \frac{\langle \underline{y}_{j+1}, \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B}. \quad (5.24)$$

The computation of the $\underline{\rho}$'s is the same as in the B -normal(ℓ, m) case when $j < m$. If $m \leq j < \theta$, the vector $\underline{\rho}_{j+1}$ can be updated using,

$$\underline{\rho}_{j+1} = \underline{rAp}_j - \sum_{i=0}^j \sigma_{i,j} \underline{\rho}_i,$$

where the quantity \underline{rAp}_j is computed the same as in the B -normal(ℓ, m) case. When $j \geq \max\{m, \theta\}$, we can begin the recursive computation of all the remainder terms. This process differs from that in the B -normal (ℓ, m) case in that additional recursions are needed to compute the remainder terms associated with the \hat{q} 's. These recursions are given by:

$$\begin{aligned} \underline{ry}_{j+1} &= \underline{rAp}_j - \sum_{i=1}^m (\eta_j)_i \underline{rq}_{j-\theta_{i-1}} - \sum_{i=1}^\kappa (\mu_j)_i \underline{r\hat{q}}_{j-\theta_{i-1}} \\ \underline{\rho}_{j+1} &= \underline{ry}_{j+1} - \sum_{i=\varphi}^j \hat{t}_{i,j} \underline{\rho}_i \end{aligned} \quad (5.25)$$

$$\underline{rq}_{j-\theta_i} = \frac{(\bar{\rho}_{j-\theta})_i}{\langle \underline{p}_{j-\theta}, \underline{p}_{j-\theta} \rangle_B} \underline{\rho}_{j-\theta} + \underline{rq}_{j-\theta-1_i}, \quad \text{for } i = 0, \dots, m-1,$$

$$\underline{r\hat{q}}_{j-\theta_i} = \frac{(\bar{\tau}_{j-\theta})_i}{\langle \underline{p}_{j-\theta}, \underline{p}_{j-\theta} \rangle_B} \underline{\rho}_{j-\theta} + \underline{r\hat{q}}_{j-\theta-1_i}, \quad \text{for } i = 0, \dots, \kappa-1.$$

An algorithm is given below to clarify the order of the computations.

ALGORITHM 5.2 CG algorithm for generalizations of B -normal(ℓ, m) matrices:

Input: $A, \underline{x}_0, \underline{r}_0, \underline{\beta}, \{\underline{\phi}_0, \dots, \underline{\phi}_{\kappa-1}\}, \theta, \varphi$.

$$\underline{p}_0 = \underline{r}_0, \\ \underline{c}p_0 = [1 \ 0 \ \dots \ 0]_{m+1}^T, \quad \underline{\rho}_0 = [1 \ 0 \ \dots \ 0]_m^T,$$

for $j = 0, \dots, \theta - 1$,

$$\underline{x}_{j+1} = \underline{x}_j + \alpha_j \underline{p}_j, \quad \underline{r}_{j+1} = \underline{r}_j - \alpha_j A \underline{p}_j, \quad \alpha_j = \frac{\langle B \underline{e}_j, \underline{p}_j \rangle}{\langle B \underline{p}_j, \underline{p}_j \rangle},$$

$$\underline{p}_{j+1} = A \underline{p}_j - \sum_{i=0}^j \sigma_{i,j} \underline{p}_i, \quad \sigma_{i,j} = \frac{\langle B A \underline{p}_j, \underline{p}_i \rangle}{\langle B \underline{p}_i, \underline{p}_i \rangle}, \\ \text{if } j < m$$

$$\underline{c}p_{j+1} = \underline{c}A \underline{p}_j - \sum_{i=0}^j \sigma_{i,j} \underline{c}p_j,$$

if $j < m - 1$

$$\underline{\rho}_{j+1} = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]_m^T,$$

if $j = m - 1$

$$\underline{\gamma} = K \setminus \underline{\beta},$$

$$\underline{\rho}_{j+1} = \left[-\frac{\gamma_0}{\gamma_m}, \dots, -\frac{\gamma_{m-1}}{\gamma_m} \right]_m^T,$$

if $j \geq m$

$$\underline{e}_j = H_{m+1,m} \underline{\rho}_j, \quad \underline{r}A \underline{p}_j = c_0 \underline{\rho}_0 + \dots + c_m \underline{\rho}_m,$$

$$\underline{\rho}_{j+1} = \underline{r}A \underline{p}_j - \sum_{i=0}^j \sigma_{i,j} \underline{\rho}_i,$$

After computing \underline{p}_θ

Solve (5.19) for $\underline{\bar{x}}_0$,

$$\underline{q}_{0_i} = \frac{(\underline{\bar{p}}_0)_i}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \underline{p}_0, \quad i = 0, \dots, m - 1,$$

$$\underline{\hat{q}}_{0_i} = \frac{(\underline{\bar{x}}_0)_i}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \underline{p}_0, \quad i = 0, \dots, \kappa - 1,$$

for $j = \theta, \dots$

$$\underline{x}_{j+1} = \underline{x}_j + \alpha_j \underline{p}_j, \quad \underline{r}_{j+1} = \underline{r}_j - \alpha_j A \underline{p}_j, \quad \alpha_j = \frac{\langle B \underline{e}_j, \underline{p}_j \rangle}{\langle B \underline{p}_j, \underline{p}_j \rangle},$$

if $j < m$

$$\underline{p}_{j+1} = A \underline{p}_j - \sum_{i=0}^j \sigma_{i,j} \underline{p}_i, \quad \sigma_{i,j} = \frac{\langle B A \underline{p}_j, \underline{p}_i \rangle}{\langle B \underline{p}_i, \underline{p}_i \rangle},$$

$$\underline{c}p_{j+1} = \underline{c}A \underline{p}_j - \sum_{i=0}^j \sigma_{i,j} \underline{c}p_j,$$

if $j < m - 1$

$$\underline{\rho}_{j+1} = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]_m^T,$$

$$\begin{aligned}
& \text{if } j = m - 1 \\
& \quad \underline{\gamma} = K \setminus \underline{\beta}, \\
& \quad \underline{\rho}_{j+1} = \left[-\frac{\gamma_0}{\gamma_m}, \dots, -\frac{\gamma_{m-1}}{\gamma_m} \right]_m^T, \\
& \text{if } j \geq m \\
& \quad \underline{y}_{j+1} = A \underline{p}_j - \left[\underline{q}_{j-\theta_0} \cdots \underline{q}_{j-\theta_{m-1}} \right] \underline{\eta}_j - \left[\hat{\underline{q}}_{j-\theta_0} \cdots \hat{\underline{q}}_{j-\theta_{\kappa-1}} \right] \underline{\mu}_j, \\
& \quad \underline{\eta}_j = [\langle A \underline{p}_j, \underline{p}_0 \rangle_B, \dots, \langle A \underline{p}_j, \underline{p}_{m-1} \rangle_B]_m^T, \\
& \quad \underline{\mu}_j = [\langle \underline{p}_j, \underline{\phi}_0 \rangle_B, \dots, \langle \underline{p}_j, \underline{\phi}_{\kappa-1} \rangle_B]_{\kappa}^T, \\
& \quad \underline{p}_{j+1} = \underline{y}_{j+1} - \sum_{i=\varphi}^j \hat{t}_{i,j} \underline{p}_i, \quad \hat{t}_{i,j} = \frac{\langle B \underline{y}_{j+1}, \underline{p}_i \rangle}{\langle B \underline{p}_i, \underline{p}_i \rangle}, \\
& \quad \underline{c}_j = H_{m+1,m} \underline{\rho}_j, \quad r A \underline{p}_j = c_0 \underline{\rho}_0 + \cdots + c_m \underline{\rho}_m, \\
& \quad r \underline{y}_{j+1} = r A \underline{p}_j - \sum_{i=1}^m (\underline{\eta}_j)_i r \underline{q}_{j-\theta_{i-1}} - \sum_{i=1}^{\kappa} (\underline{\mu}_j)_i r \hat{\underline{q}}_{j-\theta_{i-1}}, \\
& \quad \underline{\rho}_{j+1} = r \underline{y}_{j+1} - \sum_{i=\varphi}^j \hat{t}_{i,j} \underline{\rho}_i, \\
& \text{Solve (5.19) for } \bar{\underline{x}}_{j-\theta+1}, \\
& \quad \underline{q}_{j-\theta+1_i} = \frac{(\bar{\underline{p}}_{j-\theta+1})_i}{\langle \underline{p}_{j-\theta+1}, \underline{p}_{j-\theta+1} \rangle_B} \underline{p}_{j-\theta+1} + \underline{q}_{j-\theta+1_i} \quad i = 0, \dots, m-1, \\
& \quad \hat{\underline{q}}_{j-\theta+1_i} = \frac{(\bar{\underline{x}}_{j-\theta+1})_i}{\langle \underline{p}_{j-\theta+1}, \underline{p}_{j-\theta+1} \rangle_B} \underline{p}_{j-\theta+1} + \hat{\underline{q}}_{j-\theta+1_i}, \quad i = 0, \dots, \kappa-1, \\
& \quad r \underline{q}_{j-\theta+1_i} = \frac{(\bar{\underline{p}}_{j-\theta+1})_i}{\langle \underline{p}_{j-\theta+1}, \underline{p}_{j-\theta+1} \rangle_B} \underline{\rho}_{j-\theta+1} + r \underline{q}_{j-\theta+1_i}, \quad i = 0, \dots, m-1, \\
& \quad r \hat{\underline{q}}_{j-\theta+1_i} = \frac{(\bar{\underline{x}}_{j-\theta+1})_i}{\langle \underline{p}_{j-\theta+1}, \underline{p}_{j-\theta+1} \rangle_B} \underline{\rho}_{j-\theta+1} + r \hat{\underline{q}}_{j-\theta+1_i}, \quad i = 0, \dots, \kappa-1.
\end{aligned}$$

Suppose

$$A = \hat{A} + Z_r,$$

where \hat{A} is I -normal(ℓ, m), and Z_r is a rank r matrix. It follows from Corollary 4.4 that there exists polynomials p_ℓ and q_m , of degrees ℓ and m , respectively, such that

$$A^* q_m(A) - p_\ell(A) = Q_I(A), \quad \text{where,}$$

$$\text{Rank}(Q_I(A)) \leq (\ell + m + 1)r.$$

Let $B = A^* A$, and notice that

$$A^\dagger = B^{-1} A^* B = A^{-1} A^* A,$$

and

$$\begin{aligned} Q_{A^*A}(A) &= A^\dagger q_m(A) - p_\ell(A) \\ &= A^{-1} [A^* q_m(A) - p_\ell(A)] A \\ &= A^{-1} Q_I(A) A, \end{aligned}$$

and thus

$$\text{Rank}(Q_{A^*A}(A)) = \text{Rank}(Q_I(A)) \leq (\ell + m + 1)r.$$

It follows from Theorem 4.3 that a minimal residual method, $\mathcal{CG}(A^*A, A)$, can be implemented using the Algorithm 5.2.

We note here, if \hat{A} is I -self-adjoint, and $B = A^*A$, then

$$\begin{aligned} Q_{A^*A}(A) &= A^{-1} [A^* I - A] A \\ &= A^{-1} [Z_r^* - Z_r] A, \end{aligned}$$

which is A^*A skew adjoint.

One useful application of Algorithm 5.2 is on matrices that result from a discretization of a partial differential equation that is formally self-adjoint, where nonstandard boundary conditions are applied to a portion of the boundary. The next example is of this type.

Example 2: Suppose

$$A = \hat{A} + Z_r,$$

where \hat{A} is the symmetric pentadiagonal matrix that results from discretizing Poisson's equation on the unit square with homogeneous Dirchlet boundary conditions,

$$\begin{aligned} -\Delta \underline{u} &= \underline{f}, & \text{on } (0, 1) \times (0, 1), \\ \underline{u} &= \underline{0}, & \text{on } \partial\Omega, \end{aligned}$$

using a 5-point difference scheme with a uniform mesh size, $h_x = h_y = \frac{1}{11}$. The matrix Z_r is a nonsymmetric, rank 8 matrix that was chosen to correspond to perturbing the boundary conditions at a few points in the 4 corners of the domain Ω .

Note that the matrix A is I -normal(1) (self-adjoint) plus rank 8. The spectrum is plotted in Figure 5.3. Notice that the matrix is slightly indefinite, with a few eigenvalues off the real line.

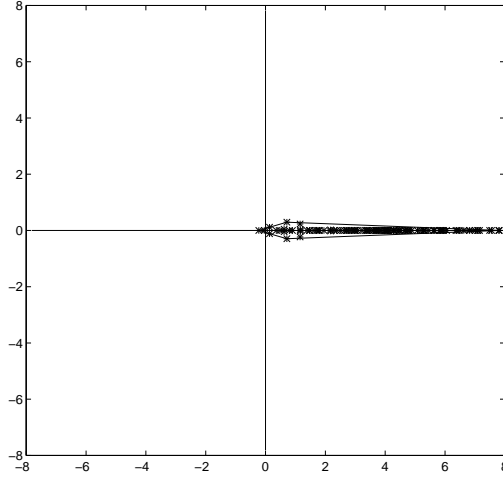


Figure 5.3: Spectrum of matrix given in Example 2.

From Corollary 4.4, we know that there exists polynomials, $p_1(A) = A$ and $q_0(A) = I$, such that

$$A^*q_0(A) - p_1(A) = A^* - A = Z_r^* - Z_r = Q_I(A), \quad \text{where } \text{Rank}(Q_I(A)) \leq 2r.$$

Algorithm 5.2 could be run on the problem using $B = I$, however, we note that this would not yield a computable algorithm, since at each step we must compute,

$$\alpha_j = \frac{\langle B\underline{e}_j, \underline{p}_j \rangle}{\langle B\underline{p}_j, \underline{p}_j \rangle},$$

where \underline{e}_j denotes the error at step j (see Section 2.3). To obtain a computable algorithm, let $B = A^*A$. From the discussion before Example 2, it follows that

$$Q_{A^*A}(A) = A^{-1}Q_I(A)A, \quad \text{where}$$

$$\text{Rank}(Q_{A^*A}(A)) \leq 2r.$$

Since \hat{A} is I -self-adjoint, a basis for the range of $Q_{A^*A}(A)$ could be obtained using (5.16). However, all that is necessary in the computations are the quantities, $\langle \underline{p}_k, \underline{\phi}_i \rangle_B$, which are computed using (5.17). A minimal residual method, $\mathcal{CG}(A^*A, A)$, is implemented using Algorithm 5.2. Figure 5.4 compares the convergence measured in the relative residual norm using Algorithm 5.2 to that using a full conjugate gradient iteration (Orthodir). As theory predicts, these two curves are identical. In addition, we plot the convergence using the Odir algorithm given in Table 2.3.

Recall that the Odir algorithm will yield the same iterates in exact arithmetic as Orthodir, if the matrix A is B -normal(1). This would be the case if we did not add the low rank matrix R_r , and $A = \hat{A}$.

We might consider how Algorithm 5.2 compares to other iterative methods for nonsymmetric matrices. The quasi-minimal residual (QMR) method of Freund and Nachtigal [7] is a method of this type. In contrast to the conjugate gradient method, $\mathcal{CG}(B, A)$ with $B = A^*A$, which minimizes the 2 norm of the residual, QMR is based upon a quasi-minimization of the residual norm. QMR can be viewed as a “variable metric” conjugate gradient method [2]. This is a conjugate gradient method where the inner product matrix B is dependent upon the initial residual. Regardless of how we view this method, for nonsymmetric matrices, QMR is not a true minimal residual method. Thus, when convergence is measured in the residual norm, QMR cannot do any better than $\mathcal{CG}(A^*A, A)$.

Figure 5.5 compares the convergence of $\mathcal{CG}(A^*A, A)$ implemented using Algorithm 5.2 to the QMR method, on Example 2. For this problem, the number of QMR iterations is comparable to the number of iterations used by Algorithm 5.2.

If we consider the main work involved in these two implementations, to be the number of matrix vector multiplications, then Algorithm 5.2 produces quite

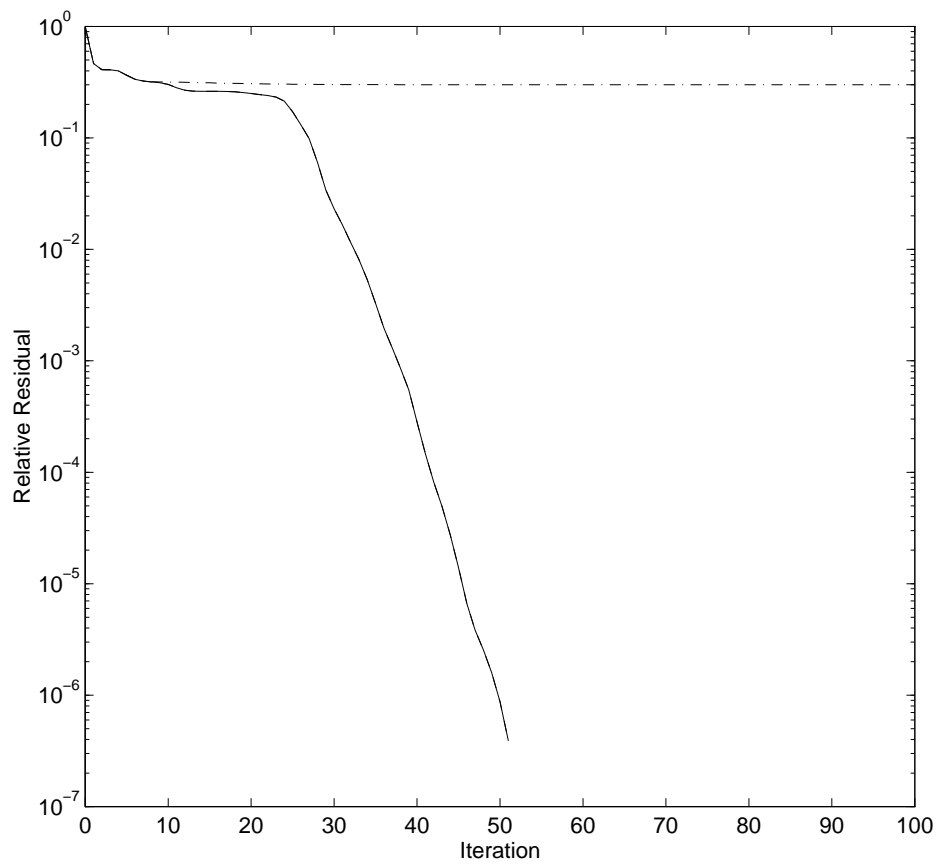


Figure 5.4. Convergence of Orthodir (solid line), Algorithm 5.2 (dashed line, plotted on top of solid line), and Odir (dashed-dotted line), on Example 2.

a savings over the QMR iteration. QMR is implemented via the nonsymmetric Lanczos method, often, using a “look-ahead” strategy [8]. This process generates two sequences of vectors, $\{\underline{v}_i\}_{i=1}^N$ and $\{\underline{w}_i\}_{i=1}^N$, such that

$$\begin{aligned} \text{sp}\{\underline{v}_1, \dots, \underline{v}_N\} &= \mathcal{K}_N(\underline{v}_1, A), \text{ and} \\ \text{sp}\{\underline{w}_1, \dots, \underline{w}_N\} &= \mathcal{K}_N(\underline{w}_1, A^*). \end{aligned}$$

This involves 2 matrix vector multiplications at each step. The computations in Algorithm 5.2 can be arranged so that only 1 matrix vector multiplication is required at each step. In Figure 5.6, we compare the main work (measured in the number of matrix vector multiplications) involved between the QMR iteration and Algorithm 5.2, on Example 2.

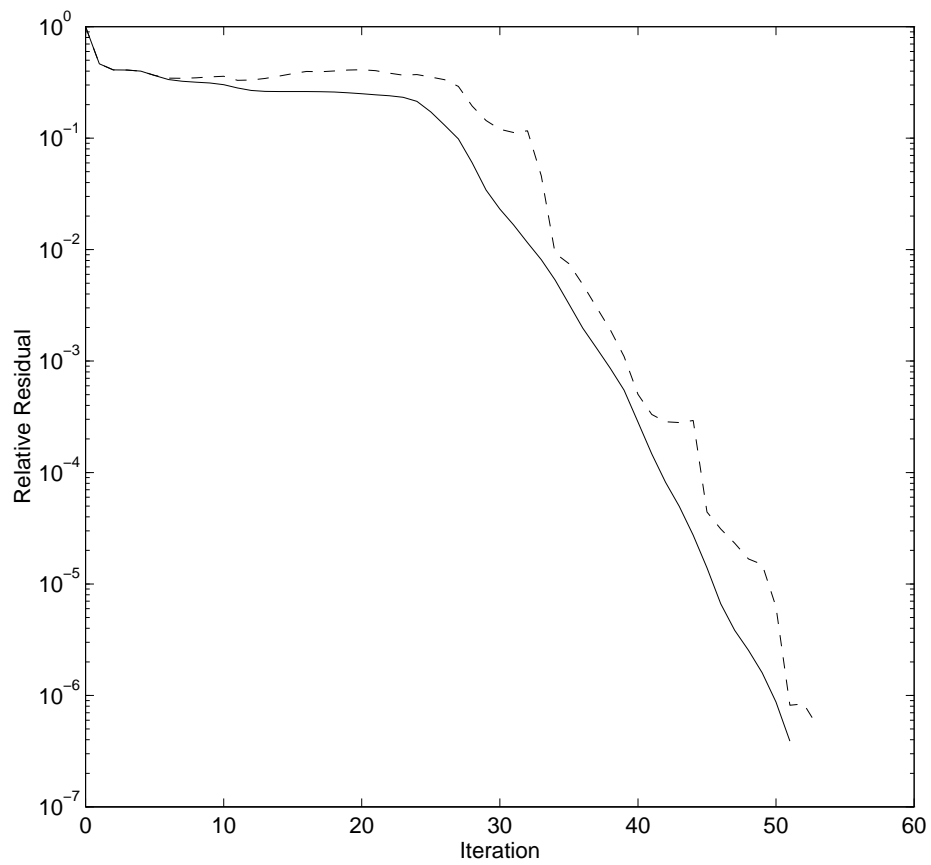


Figure 5.5. Convergence of Algorithm 5.2 (solid line), and QMR (dashed line), on Example 2.

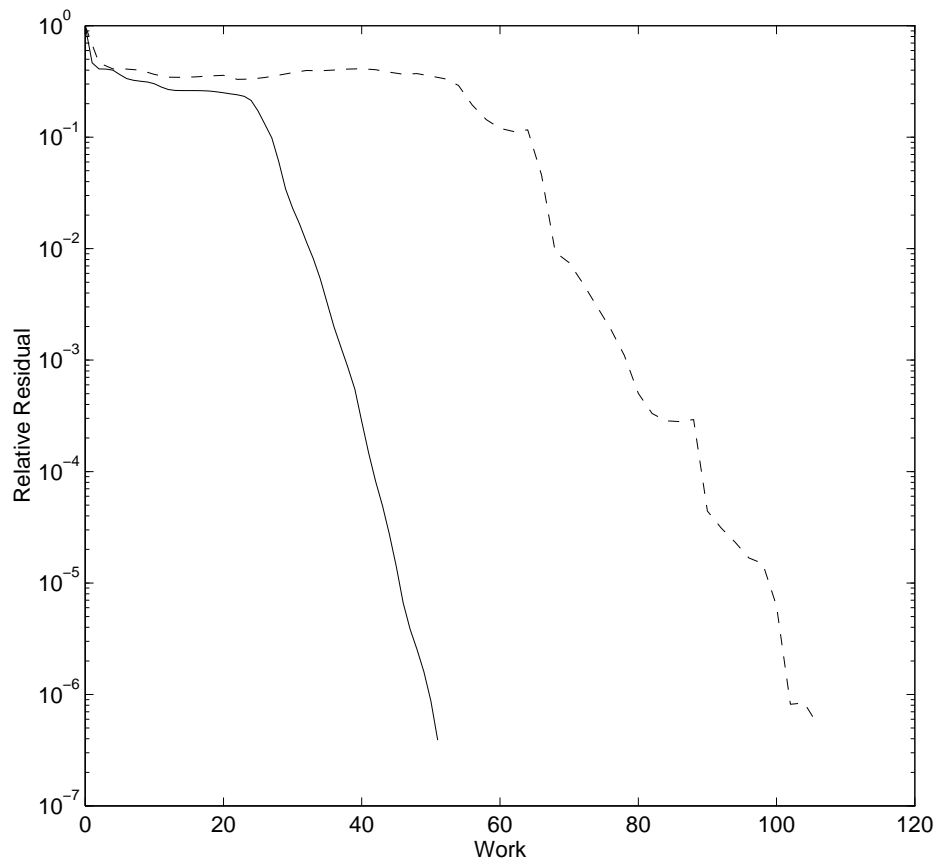


Figure 5.6. Work in matrix vector multiplications for Algorithm 5.2 (solid line), and QMR (dashed line), on Example 2.

Example 3: The next example is the matrix

$$A = \hat{A} + Z_r,$$

where \hat{A} is the shifted unitary matrix in Example 1, and Z_r is a rank 2 matrix.

Figure 5.7 plots the eigenvalues of this matrix.

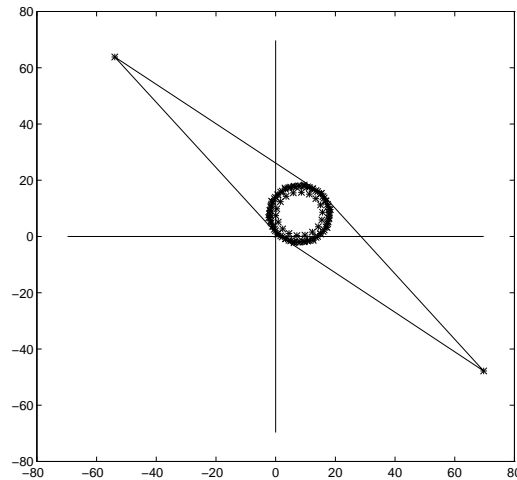


Figure 5.7: Spectrum of problem given in Example 3.

Again, we will take $B = A^*A$, and run Algorithm 5.2. We compare the convergence of Algorithm 5.2 to the Orthodir algorithm and the QMR iteration in Figure 5.8. Note that the convergence curves of Orthodir and Algorithm 5.2 are plotted on top of one another, as expected. In Figure 5.9, we compare the main work (measured in the number of matrix vector multiplications) between Algorithm 5.2 and the QMR iteration, on the problem given in Example 3. Notice that the work involved using the QMR iteration is more than double the amount used by Algorithm 5.2.

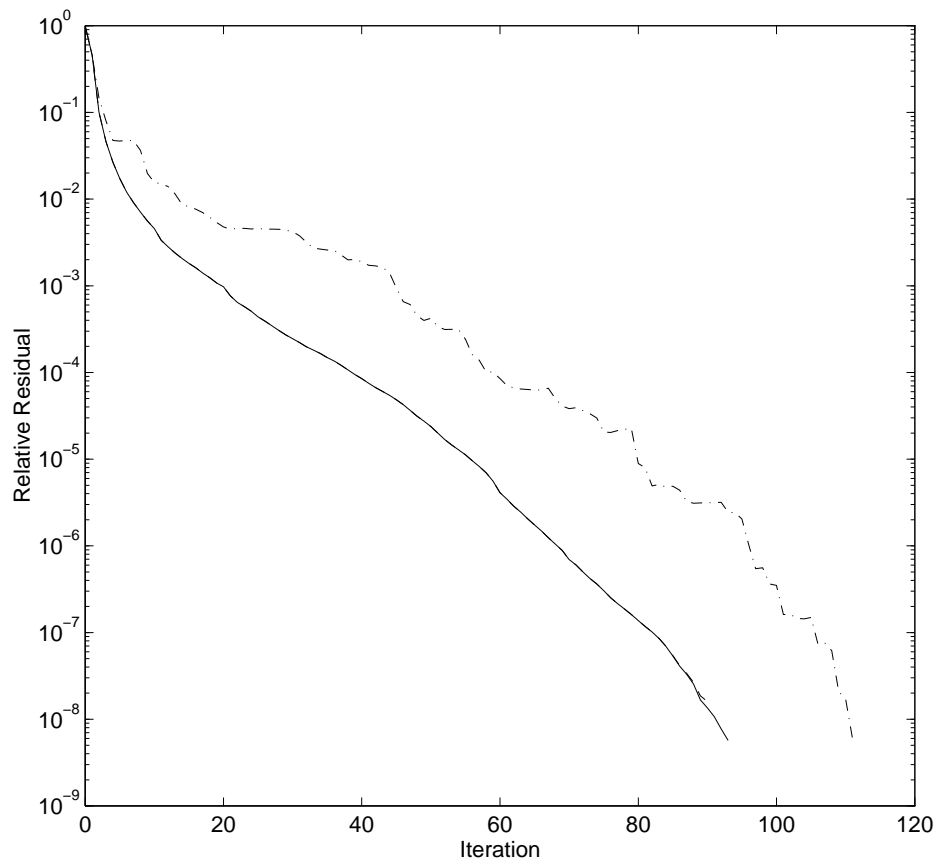


Figure 5.8. Convergence of Orthodir (solid line), Algorithm 5.2 (dashed line, plotted on top of solid line), and QMR (dashed-dotted line), on Example 3.

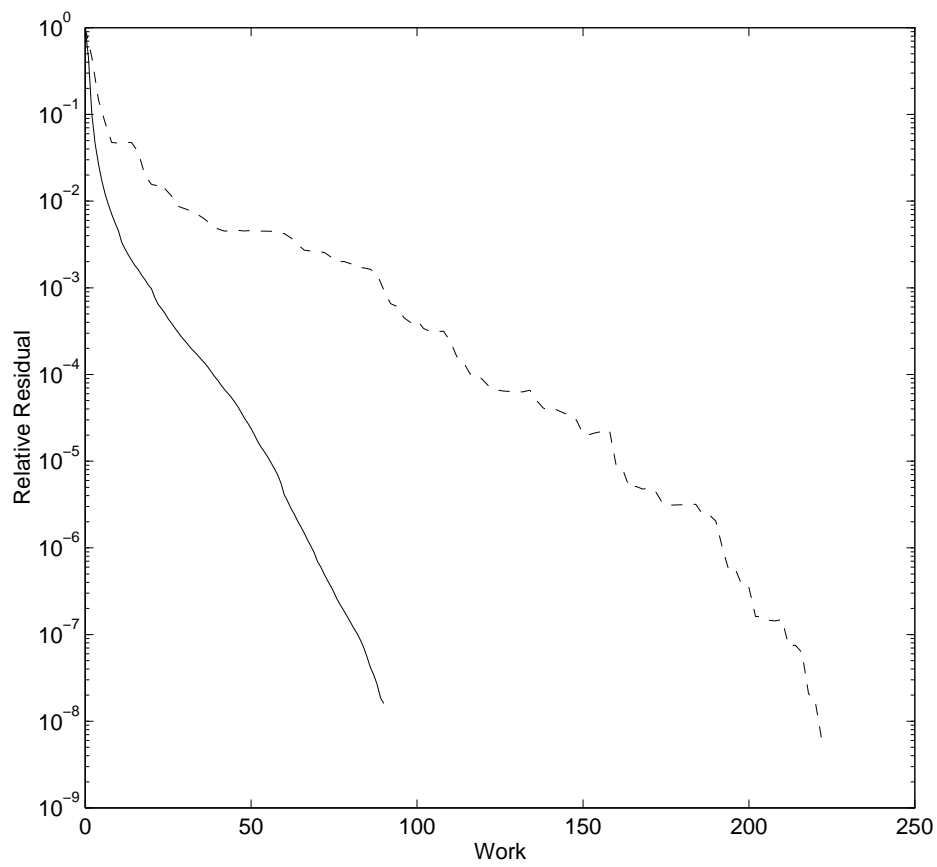


Figure 5.9. Work in matrix vector multiplications of Algorithm 5.2 (solid line), and QMR (dashed line), on Example 3.

5.4 Concluding Remarks

In this chapter, we presented general algorithms for implementing the conjugate gradient method using the multiple recursions given in (4.21) and (4.35). We note here, that in certain special cases, these algorithms simplify. For example, if A is self-adjoint plus low rank, the polynomial q_m is a constant polynomial. Since no remainder results upon division by a constant polynomial, the vector $\underline{\rho}_k$ is the zero vector. The multiple recursion involves only one set of auxiliary vectors, the $\underline{\hat{q}}_{k_i}$'s.

Only small numerical examples have been run to validate the theory. No attempt has been made to analyze the effect of numerical round off on the iteration. It may be possible in some cases, for example, if A is self-adjoint plus low rank, to extend the analysis done by Greenbaum ([12]).

6. Necessary Conditions

6.1 Introduction

In this chapter, we will be concerned with determining necessary conditions on the matrix A required for multiple recursions of the form introduced in Chapter 4, to yield a B -orthogonal basis for $\mathcal{K}_d(\underline{x}_0, A)$, for every \underline{x}_0 . This is a difficult problem, and we will confine our analysis to a few special cases of recursions of this form.

Sufficient conditions on the matrix A have already been established for multiple recursions of the forms given in (4.21) and (4.35). These conditions are for A to be either B -normal(ℓ, m), or a generalization of a B -normal(ℓ, m) matrix (see Section 4.4). We recall that for these matrices, in the absence of breakdown, a single (s, t) -recursion could also be used to construct this basis.

Since the tools we will use in this analysis are based upon the single (s, t) -recursion, we will begin the next section with some results pertaining to the likelihood of a breakdown occurring at some step in the (s, t) -iteration.

Based upon the formulation of the multiple recursions in Chapter 4, for B -normal(ℓ, m) and generalizations of B -normal(ℓ, m) matrices, we will define a general form of multiple recursion. The relationship between breakdown in the single (s, t) -recursion and the corresponding multiple recursion will be discussed. In particular, it will be shown that for a restricted subset of multiple recursions of this form, breakdown in the corresponding single (s, t) -recursion can be limited to a set of initial vectors \underline{p}_0 of measure zero in C^N . These are the multiple recursions that will be analyzed in this chapter. After developing some tools to use in this analysis, the remainder of the chapter will contain the main proof establishing necessary

conditions.

6.2 Preliminaries

We begin with some preliminary material that will be used throughout this chapter.

LEMMA 6.1 If p is a complex nonzero multivariate polynomial, then $p(x_1, \dots, x_N) \neq 0$ for almost every $\underline{x} = [x_1, \dots, x_N]^T \in \mathcal{C}^N$.

Proof: The proof can be found in [20]. \square

In other words, this lemma says that if a polynomial is not the zero polynomial, then the set of $\underline{x} \in \mathcal{C}^N$ for which the polynomial can be zero, is a set of measure zero in \mathcal{C}^N .

Consider the multivariate rational function

$$f(x_1, \dots, x_N) = \frac{p(x_1, \dots, x_N)}{q(x_1, \dots, x_N)}.$$

Notice that if f is finite anywhere, then q cannot be the zero polynomial, and it follows that f is finite almost everywhere. Further, if f is finite, it can either be zero on a set of vectors $\underline{x} \in \mathcal{C}^N$ of measure zero, or it can be identically zero.

Given an initial vector \underline{p}_0 , recall that an ascending B -orthogonal basis $\{\underline{p}_j\}_{j=0}^{d-1}$ for $\mathcal{K}_d(\underline{p}_0, A)$ is unique up to scale. In some cases, this basis may be computed with some form of a short recursion. The same basis can always be computed using a full Gram-Schmidt process generalized to the B -inner product (2.8).

In [5], Faber and Manteuffel proved that for $i \leq d(A)$, that \underline{p}_i is a continuous function of \underline{p}_0 , and that

$$\|\underline{p}_i\| \leq \|A\| \|\underline{p}_{i-1}\|.$$

A similar line of proof can be used to show each component of \underline{p}_i is a bounded rational function of the components of \underline{p}_0 .

LEMMA 6.2 Suppose \underline{p}_i is computed as in (2.8). Let $\underline{p}_0 = [z_1, \dots, z_N]^T$, where $z_j \in \mathcal{C}$. Then for all $i \leq d(A)$, each component, $(\underline{p}_i)_j$ of \underline{p}_i is a bounded rational function of (z_1, \dots, z_N) .

Proof: Each component of \underline{p}_0 , being a linear function of it's real and imaginary parts, is rational. Next, consider the computation for \underline{p}_1 ,

$$\underline{p}_1 = A\underline{p}_0 - \frac{\langle A\underline{p}_0, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \underline{p}_0.$$

Denoting the (j, k) 'th element of the matrix A as $(a + bi)_{j,k}$, we see that the j 'th component of $A\underline{p}_0$ is given by

$$(A\underline{p}_0)_j = \sum_{k=1}^N (a + bi)_{j,k} z_k.$$

This is a first degree polynomial in the components of \underline{p}_0 . It follows that the inner products, $\langle A\underline{p}_0, \underline{p}_0 \rangle_B$ and $\langle \underline{p}_0, \underline{p}_0 \rangle_B$, are each the sum of terms involving the products of two first degree polynomials, yielding second degree polynomials in (z_1, \dots, z_N) . By combining this information it follows that the j 'th component of \underline{p}_1 can be written as,

$$(\underline{p}_1)_j = \frac{m_3^{(j)}(z_1, \dots, z_N)}{q_2^{(j)}(z_1, \dots, z_N)},$$

for some polynomials $m_3^{(j)}$ and $q_2^{(j)}$ of degrees three and two, respectively. From the proof in [5], we know that $(\underline{p}_1)_j$ is finite.

Suppose that for $i < d(A)$, and that, for $\ell \leq i$, $\underline{p}_\ell \neq \underline{0}$ and

$$p_\ell^{(j)} = \frac{m_{t_\ell}^{(j)}(z_1, \dots, z_N)}{q_{s_\ell}^{(j)}(z_1, \dots, z_N)} < \infty,$$

for some polynomials $m_{t_\ell}^{(j)}$ and $q_{s_\ell}^{(j)}$ of degrees t_ℓ and s_ℓ , respectively. Consider

$$\underline{p}_{i+1} = A\underline{p}_i - \sum_{\ell=0}^i \sigma_{\ell,i} \underline{p}_\ell.$$

The j 'th component of $A\underline{p}_i$ is given by,

$$\begin{aligned} (A\underline{p}_i)_j &= \sum_{k=1}^N (a + bi)_{j,k} (\underline{p}_i)_k \\ &= \sum_{k=1}^N (a + bi)_{j,k} \frac{m_{t_i}^{(k)}(z_1, \dots, z_N)}{q_{s_i}^{(k)}(z_1, \dots, z_N)}. \end{aligned}$$

Since this is a sum of bounded rational functions, it is also bounded and rational. Notice that inner products of the form $\langle A\underline{p}_i, \underline{p}_\ell \rangle_B$ and $\langle \underline{p}_\ell, \underline{p}_\ell \rangle_B$ are bounded and rational since they involve sums and products of bounded rational functions. The terms $\sigma_{\ell,i} = \frac{\langle A\underline{p}_i, \underline{p}_\ell \rangle_B}{\langle \underline{p}_\ell, \underline{p}_\ell \rangle_B}$, for $\ell = 0, \dots, i$, are quotients of rational functions, thus rational. Since $\underline{p}_\ell \neq \underline{0}$, and B is Hermitian positive definite, $\langle \underline{p}_\ell, \underline{p}_\ell \rangle_B \neq 0$, and it follows that the terms $\sigma_{\ell,i}$ are also bounded. Therefore, the j 'th component of \underline{p}_{i+1} ,

$$(\underline{p}_{i+1})_j = (A\underline{p}_i)_j - \sum_{\ell=0}^i \sigma_{\ell,i} (\underline{p}_\ell)_j,$$

involves sums and products of bounded rational functions, yielding a bounded rational function. \square

Suppose that A is B -normal(ℓ, m), or a generalization of a B -normal(ℓ, m) matrix (see Sections 4.3 and 4.4). In Chapter 4, we saw that a single (s, t) -recursion could be used to construct a B -orthogonal basis for $\mathcal{K}_d(\underline{r}_0, A)$, if the corresponding systems given in (4.13) and (4.26) were consistent at each step. We note here, that the elements of the system matrices in (4.13) and (4.26) are of the form,

$$\langle t_1(\underline{p}_0), t_2(\underline{p}_0) \rangle,$$

where t_1 and t_2 are bounded rational functions of \underline{p}_0 . Since these inner products involve sums and products of bounded rational functions, they are bounded and rational. The determinant of a matrix involves sums and products its elements. Thus, the determinants of the matrices in (4.13) and (4.26) are bounded rational functions of the components of \underline{p}_0 . From the discussion following Lemma 6.1 we see

that these determinants can either be zero for every \underline{p}_0 , or only for a set of \underline{p}_0 of measure zero. This means that, for B -normal(ℓ, m) matrices, and generalizations of B -normal(ℓ, m) matrices, breakdown in the single (s, t) -recursion occurs either for every \underline{p}_0 , or only for a measure zero set of initial vectors \underline{p}_0 . It would be desirable to rule out the possibility of breakdown for every \underline{p}_0 . If this could be done, then we could say the likelihood of a breakdown occurring at any step in the (s, t) -recursion in exact arithmetic is rare.

Suppose $d(\underline{x}_0, A) = N$, and recall that for any matrix A , a B -orthogonal basis for $\mathcal{K}_d(\underline{x}_0, A)$ can be constructed using a full Gram-Schmidt process (2.8), which yields the matrix equation

$$AP_N = P_N H_N.$$

If A is B -normal(ℓ, m), the Hessenberg matrix H_N can be decomposed as,

$$H_N = T + U,$$

where T is a banded upper Hessenberg matrix with 1's on the subdiagonal, and an upper bandwidth of $\max\{0, \ell - m + 1\}$, and U is an upper triangular, factorable matrix, with components:

$$u_{i,j} = \begin{cases} \frac{\langle \underline{\eta}_j, \underline{\rho}_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j \geq i \\ 0 & j < i \end{cases},$$

where $\underline{\eta}_j$ and $\underline{\rho}_i$ have length m . From Section 4.3, notice that the elements of both T and U are bounded rational functions of \underline{p}_0 . This decomposition of H_N results in a multiple recursion (4.21) that involves m auxiliary vectors at each step.

If A is a generalization of a B -normal(ℓ, m) matrix (see Section 4.4), the Hessenberg matrix can be written as:

$$\begin{aligned} H_N &= T + U, \\ U &= U_1 + U_2, \end{aligned}$$

where T is banded upper Hessenberg, with 1's on the subdiagonal, and an upper bandwidth of $\max\{0, \ell - m + 1\}$. The elements of U_1 and U_2 are given by:

$$u_{i,j}^{(1)} = \begin{cases} \frac{\langle \eta_j, \underline{p}_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j \geq i \\ 0 & j < i \end{cases}, \quad u_{i,j}^{(2)} = \begin{cases} \frac{\langle \underline{\mu}_j, \tau_i \rangle}{\langle \underline{p}_i, \underline{p}_i \rangle_B} & j \geq i \\ 0 & j < i \end{cases},$$

respectively. Again, the components of T , U_1 , and U_2 are bounded rational functions of \underline{p}_0 . By denoting the combined vectors,

$$\underline{w}_j = \begin{pmatrix} \eta_j \\ \underline{\mu}_j \end{pmatrix}, \quad \text{and} \quad \underline{v}_i = \frac{1}{\langle \underline{p}_i, \underline{p}_i \rangle_B} \begin{pmatrix} \rho_i \\ \tau_i \end{pmatrix},$$

of length $m + \kappa$, where κ is the rank of $Q_B(A)$, the entries of the matrix U can then be written as,

$$u_{i,j} = \begin{cases} \langle \underline{w}_j, \underline{v}_i \rangle & j \geq i \\ 0 & j < i \end{cases}.$$

This decomposition of H_N led to a multiple recursion (4.35) involving $m + \kappa$ auxiliary vectors at each step.

Based on these recursions for B -normal(ℓ, m) and generalizations of B -normal(ℓ, m) matrices, we define a general form of recursion. Suppose A is such that the elements of the Hessenberg matrix H_N , resulting from a full Gram-Schmidt process, simplify above some upper bandwidth b :

$$H_N = \begin{bmatrix} \frac{\langle A\underline{p}_0, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \cdots & \frac{\langle A\underline{p}_{\ell-1}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \langle \underline{w}_b, \underline{v}_0 \rangle & \cdots & \langle \underline{w}_{N-1}, \underline{v}_0 \rangle \\ 1 & \frac{\langle A\underline{p}_1, \underline{p}_1 \rangle_B}{\langle \underline{p}_1, \underline{p}_1 \rangle_B} & & \frac{\langle A\underline{p}_b, \underline{p}_1 \rangle_B}{\langle \underline{p}_1, \underline{p}_1 \rangle_B} & \cdots & \vdots \\ & \ddots & \ddots & \ddots & \ddots & \langle \underline{w}_{N-1}, \underline{v}_{N-b-1} \rangle \\ & & \ddots & \ddots & \ddots & \frac{\langle A\underline{p}_{N-1}, \underline{p}_{N-b} \rangle_B}{\langle \underline{p}_{N-b}, \underline{p}_{N-b} \rangle_B} \\ & & & \ddots & \ddots & \vdots \\ & & & & \ddots & \frac{\langle A\underline{p}_{N-1}, \underline{p}_{N-1} \rangle_B}{\langle \underline{p}_{N-1}, \underline{p}_{N-1} \rangle_B} \\ & & & & 1 & \end{bmatrix}. \quad (6.1)$$

The matrix H_N can be written as:

$$H_N = T + U,$$

where the matrix T is a banded upper Hessenberg matrix, with upper bandwidth of b , and a subdiagonal consisting of all 1's. The matrix U is upper triangular and factorable, with elements of the form,

$$u_{i,j} = \begin{cases} \langle \underline{w}_j, \underline{v}_i \rangle, & j \geq i, \\ 0, & j < i, \end{cases}. \quad (6.2)$$

The vectors \underline{w}_j and \underline{v}_i have length t and are each bounded rational functions of \underline{p}_0 . The matrix U could be further decomposed as,

$$U = U_1 + \cdots + U_n,$$

where \underline{w}_j and \underline{v}_i could be written in terms of the combined vectors from the individual matrices U_k ,

$$\underline{w}_j = \begin{pmatrix} \underline{w}_{1j} \\ \vdots \\ \underline{w}_{nj} \end{pmatrix}, \quad \text{and} \quad \underline{v}_i = \begin{pmatrix} \underline{v}_{1i} \\ \vdots \\ \underline{v}_{ni} \end{pmatrix}. \quad (6.3)$$

Notice that the elements of T and U are all bounded rational functions of \underline{p}_0 . Analogous to the recursions derived for the B -normal(ℓ, m) case, and the generalizations of the B -normal(ℓ, m) case, this decomposition yields a set of multiple recursions that involves t auxiliary vectors:

$$\begin{aligned} \underline{p}_{j+1} &= A\underline{p}_j + \sum_{i=j-(b-1)}^j t_{i,j} \underline{p}_i - [\underline{q}_{j0} \cdots \underline{q}_{jt-1}] \underline{w}_j, \\ \underline{q}_{j+1_i} &= (\underline{v}_{j+1})_i \underline{p}_{j+1} + \underline{q}_{ji}, \quad \text{for } i = 0, \dots, t-1. \end{aligned} \quad (6.4)$$

We denote a multiple recursion of this form as $MR(b, t)$, where b is the upper bandwidth of the Hessenberg matrix T , and t denotes the length of the vectors \underline{w}_j and \underline{v}_i given in (6.3).

This form of multiple recursion includes the recursions for both the B -normal(ℓ, m) case, and the generalizations of the B -normal(ℓ, m) case. Notice that

in both cases, $b = \max\{0, \ell - m + 1\}$. In the B -normal(ℓ, m) case, $t = m$, which is the length of the vectors, $\underline{\eta}_j$ (4.16), whereas, for the generalizations of B -normal(ℓ, m) matrices, $t = m + \kappa$, which is the length of the combined vectors,

$$\begin{pmatrix} \underline{\eta}_j \\ \underline{\mu}_j \end{pmatrix},$$

given in (4.29) and (4.31).

DEFINITION 6.1 A matrix A is in the class $\mathcal{CG}_{\text{MR}}(b, t)$ if for every \underline{x}_0 , a B -orthogonal basis can be constructed using a multiple recursion of the form $MR(b, t)$ given in (6.4).

The sufficient conditions given in Chapter 4 can be restated in terms of this definition. We state these in the following Theorem.

THEOREM 6.3 Sufficient Conditions for $A \in \mathcal{CG}_{\text{MR}}(b, t)$:

- (1) If A is B -normal(ℓ, m), then $A \in \mathcal{CG}_{\text{MR}}(b, t)$, where $b = \max\{0, \ell - m + 1\}$, and $t = m$.
- (2) If there exists polynomials p_ℓ and q_m of degrees ℓ and m , respectively, satisfying

$$\begin{aligned} A^\dagger q_m(A) - p_\ell(A) &= Q_B(A), \quad \text{where} \\ \text{Rank}(Q_B(A)) &= \kappa, \end{aligned}$$

then $A \in \mathcal{CG}_{\text{MR}}(b, t)$, with $b = \max\{0, \ell - m + 1\}$, and $t = m + \kappa$.

Proof: See Theorems 4.2 and 4.3 in Chapter 4. \square

To determine if there are any other matrices in the class $\mathcal{CG}_{\text{MR}}(b, t)$, we must establish whether or not the sufficient conditions are also necessary. The proof establishing necessary conditions for $A \in \mathcal{CG}_{\text{MR}}(b, t)$, will be based upon tools developed for analyzing a corresponding single (s, t) -recursion. If $A \in \mathcal{CG}_{\text{MR}}(b, t)$,

a B -orthogonal basis for $\mathcal{K}_d(\underline{r}_0, A)$ can be constructed using a multiple recursion $MR(b, t)$. We will define a corresponding single (s, t) -recursion, and show what conditions must be satisfied in order for this basis to be constructed using this (s, t) -recursion.

Recall that in order for a single (s, t) -recursion ,

$$\underline{p}_{j+1} = A\underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} A\underline{p}_k - \sum_{i=j-s}^j \sigma_{i,j} \underline{p}_i, \quad (6.5)$$

to yield a B -orthogonal basis for $\mathcal{K}_d(\underline{r}_0, A)$, for every $0 \leq j \leq d(\underline{p}_0, A) - 2$, there exists $\{\beta_{k,j}\}_{k=j-t}^{j-1}$ such that,

$$\sigma_{i,j} = \frac{\langle A\underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} A\underline{p}_k, \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B} = 0, \quad \text{for } i = 0, \dots, j - s - 1.$$

This condition can be rewritten as:

$$\begin{bmatrix} \frac{\langle A\underline{p}_{j-t}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \dots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ \vdots & & \vdots \\ \frac{\langle A\underline{p}_{j-t}, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} & \dots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} \\ \vdots & & \vdots \\ \frac{\langle A\underline{p}_{j-t}, \underline{p}_{j-s-1} \rangle_B}{\langle \underline{p}_{j-s-1}, \underline{p}_{j-s-1} \rangle_B} & \dots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_{j-s-1} \rangle_B}{\langle \underline{p}_{j-s-1}, \underline{p}_{j-s-1} \rangle_B} \end{bmatrix} \begin{pmatrix} \beta_{j-t,j} \\ \vdots \\ \beta_{j-1,j} \end{pmatrix} = - \begin{pmatrix} \frac{\langle A\underline{p}_j, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ \vdots \\ \frac{\langle A\underline{p}_j, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} \\ \vdots \\ \frac{\langle A\underline{p}_j, \underline{p}_{j-s-1} \rangle_B}{\langle \underline{p}_{j-s-1}, \underline{p}_{j-s-1} \rangle_B} \end{pmatrix}. \quad (6.6)$$

If $A \in \mathcal{CG}_{\text{MR}}(b, t)$, the matrix H_N simplifies as in (6.1). So when $j \geq i + b$, the elements of H_N , $\frac{\langle A\underline{p}_j, \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B} = \langle \underline{w}_j, \underline{v}_i \rangle$, where the vectors \underline{w}_j , and \underline{v}_i are of length t . Recall the matrix equation after constructing \underline{p}_{j+1} using (2.8),

$$A\underline{p}_{j+1} = P_{j+2} H_{j+2, j+1}.$$

When $j = b + t$, the upper right $1 \times (t + 1)$ corner of $H_{j+2, j+1}$ can be written as:

$$\left[\frac{\langle A\underline{p}_{j-t}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \dots \frac{\langle A\underline{p}_{j-1}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \frac{\langle A\underline{p}_j, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \right]$$

$$= \left[\langle \underline{w}_{j-t}, \underline{v}_0 \rangle \cdots \langle \underline{w}_{j-1}, \underline{v}_0 \rangle \langle \underline{w}_j, \underline{v}_0 \rangle \right].$$

At each additional step, another row in the upper right corner of $H_{j+2,j+1}$ simplifies.

When $j = 2t + b - 1$, the upper right $t \times (t + 1)$ corner of $H_{j+2,j+1}$ is given by:

$$\begin{aligned} & \begin{bmatrix} \frac{\langle A\underline{p}_{j-t}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \cdots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \frac{\langle A\underline{p}_j, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ \vdots & & \vdots & \vdots \\ \frac{\langle A\underline{p}_{j-t}, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} & \cdots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} & \frac{\langle A\underline{p}_j, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} \end{bmatrix} \\ &= \begin{bmatrix} \langle \underline{w}_{j-t}, \underline{v}_0 \rangle & \cdots & \langle \underline{w}_{j-1}, \underline{v}_0 \rangle & \langle \underline{w}_j, \underline{v}_0 \rangle \\ \vdots & & \vdots & \vdots \\ \langle \underline{w}_{j-t}, \underline{v}_{t-1} \rangle & \cdots & \langle \underline{w}_{j-1}, \underline{v}_{t-1} \rangle & \langle \underline{w}_j, \underline{v}_{t-1} \rangle \end{bmatrix}. \end{aligned}$$

In general, if $j \geq b + t$, we have,

$$\begin{aligned} & \begin{bmatrix} \frac{\langle A\underline{p}_{j-t}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \cdots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \frac{\langle A\underline{p}_j, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ \vdots & & \vdots & \vdots \\ \frac{\langle A\underline{p}_{j-t}, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} & \cdots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} & \frac{\langle A\underline{p}_j, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} \\ \vdots & & \vdots & \vdots \\ \frac{\langle A\underline{p}_{j-t}, \underline{p}_{j-s-1} \rangle_B}{\langle \underline{p}_{j-s-1}, \underline{p}_{j-s-1} \rangle_B} & \cdots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_{j-s-1} \rangle_B}{\langle \underline{p}_{j-s-1}, \underline{p}_{j-s-1} \rangle_B} & \frac{\langle A\underline{p}_j, \underline{p}_{j-s-1} \rangle_B}{\langle \underline{p}_{j-s-1}, \underline{p}_{j-s-1} \rangle_B} \end{bmatrix} \tag{6.7} \\ &= \begin{bmatrix} \langle \underline{w}_{j-t}, \underline{v}_0 \rangle & \cdots & \langle \underline{w}_{j-1}, \underline{v}_0 \rangle & \langle \underline{w}_j, \underline{v}_0 \rangle \\ \vdots & & \vdots & \vdots \\ \langle \underline{w}_{j-t}, \underline{v}_{t-1} \rangle & \cdots & \langle \underline{w}_{j-1}, \underline{v}_{t-1} \rangle & \langle \underline{w}_j, \underline{v}_{t-1} \rangle \\ \vdots & & \vdots & \vdots \\ \langle \underline{w}_{j-t}, \underline{v}_{j-s-1} \rangle & \cdots & \langle \underline{w}_{j-1}, \underline{v}_{j-s-1} \rangle & \langle \underline{w}_j, \underline{v}_{j-s-1} \rangle \end{bmatrix}, \end{aligned}$$

where,

$$s = b + t - 1.$$

Notice that if

$$\begin{bmatrix} | & & | \\ \underline{w}_{j-t} & \cdots & \underline{w}_{j-1} \\ | & & | \end{bmatrix} \in \mathcal{C}^{t \times t} \quad (6.8)$$

is nonsingular, then there always exists $\{\beta_{k,j}\}_{k=j-t}^{j-1}$ such that

$$\begin{bmatrix} | & & | \\ \underline{w}_{j-t} & \cdots & \underline{w}_{j-1} \\ | & & | \end{bmatrix} \begin{pmatrix} \beta_{j-t,j} \\ \vdots \\ \beta_{j-1,j} \end{pmatrix} = - \begin{pmatrix} | \\ \underline{w}_j \\ | \end{pmatrix}.$$

By multiplying both sides on the left by

$$\begin{bmatrix} - & \bar{v}_0 & - \\ & \vdots & \\ - & \bar{v}_{t-1} & - \\ & \vdots & \\ - & \bar{v}_{j-s-1} & - \end{bmatrix},$$

and using (6.7), we obtain,

$$\begin{bmatrix} \frac{\langle A\underline{p}_{j-t}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \cdots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ \vdots & & \vdots \\ \frac{\langle A\underline{p}_{j-t}, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} & \cdots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} \\ \vdots & & \vdots \\ \frac{\langle A\underline{p}_{j-t}, \underline{p}_{j-s-1} \rangle_B}{\langle \underline{p}_{j-s-1}, \underline{p}_{j-s-1} \rangle_B} & \cdots & \frac{\langle A\underline{p}_{j-1}, \underline{p}_{j-s-1} \rangle_B}{\langle \underline{p}_{j-s-1}, \underline{p}_{j-s-1} \rangle_B} \end{bmatrix} \begin{pmatrix} \beta_{j-t,j} \\ \vdots \\ \beta_{j-1,j} \end{pmatrix} = - \begin{pmatrix} \frac{\langle A\underline{p}_j, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} \\ \vdots \\ \frac{\langle A\underline{p}_j, \underline{p}_{t-1} \rangle_B}{\langle \underline{p}_{t-1}, \underline{p}_{t-1} \rangle_B} \\ \vdots \\ \frac{\langle A\underline{p}_j, \underline{p}_{j-s-1} \rangle_B}{\langle \underline{p}_{j-s-1}, \underline{p}_{j-s-1} \rangle_B} \end{pmatrix}.$$

This is the condition that must be satisfied in order that an (s, t) -recursion will yield a B -orthogonal basis. Notice from (6.5), that until $j = b + t = s + 1$, breakdown is not possible in the (s, t) -recursion. Therefore, it follows that if $A \in \mathcal{CG}_{\text{MR}}(b, t)$, and if (6.8) is nonsingular for every $j \geq b + t$, then an (s, t) -recursion, with $s = b + t - 1$, can also be used to construct a B -orthogonal basis.

LEMMA 6.4 Suppose $A \in \mathcal{CG}_{\text{MR}}(b, t)$. The Hessenberg matrix H_N that results after N steps of a full Gram-Schmidt process can be written as in (6.1). H_N can be decomposed as

$$H_N = T + U,$$

where T is upper Hessenberg with upper bandwidth b , and U is upper triangular with elements given by (6.2). If

$$\begin{bmatrix} | & & | \\ \underline{w}_{j-t} & \cdots & \underline{w}_{j-1} \\ | & & | \end{bmatrix} \in \mathcal{C}^{t \times t} \quad (6.9)$$

is nonsingular for every $j \geq b+t$, then a single (s, t) -recursion (6.5), with $s = b+t-1$, can also be used to construct a B -orthogonal basis for $\mathcal{K}_d(\underline{x}_0, A)$.

Proof: See above discussion. \square

Since $\underline{w}_{j-t}(\underline{p}_0), \dots, \underline{w}_{j-1}(\underline{p}_0)$, are each bounded rational functions of \underline{p}_0 , it follows that the determinant of the matrix in (6.9) is also a bounded rational function of \underline{p}_0 . Recall that this can either be zero for every \underline{p}_0 , or for a set of \underline{p}_0 of measure zero. This means that if $A \in \mathcal{CG}_{\text{MR}}(b, t)$, breakdown in the corresponding single (s, t) -recursion, with $s = b+t-1$, either happens for every \underline{p}_0 , or only for at most a set of \underline{p}_0 of measure zero. We will show that if $A \in \mathcal{CG}_{\text{MR}}(b, t)$, with $t \leq 1$, breakdown can be limited to a set of initial vectors \underline{p}_0 of measure zero.

Consider the class $\mathcal{CG}_{\text{MR}}(b, t)$, with $t \leq 1$. If $t = 0$, the multiple recursion $MR(b, t)$ given in (6.4) becomes a single recursion. If $b = 0$, $MR(0, 0)$ is given by,

$$\underline{p}_{j+1} = A\underline{p}_j.$$

Notice that in order for this recursion to yield a B -orthogonal basis,

$$\langle A\underline{p}_0, \underline{p}_0 \rangle_B = 0, \quad \text{for every } \underline{p}_0.$$

By choosing \underline{p}_0 to be any eigenvector of A , it follows that the only way this condition can be met, is if all the eigenvalues of A are zero. This is impossible for nonsingular

A. If $b > 0$, $MR(b, 0)$ is given by,

$$\underline{p}_{j+1} = A\underline{p}_j + \sum_{i=j-(b-1)}^j t_{i,j} \underline{p}_i.$$

Notice that this is an (s, t) -recursion with $s = b - 1$ and $t = 0$. Breakdown is not possible here, in fact, this recursion is the same as the $(s + 2)$ -term recursion (2.11) with $s = b - 1$, studied by Faber and Manteuffel in [5]. They showed that the only matrices for which an orthogonal basis could be constructed with this recursion are B -normal $(b - 1)$ matrices, or matrices with $d(A) \leq b + 1$ (see Theorem 2.1).

The next theorem shows that if $A \in \mathcal{CG}_{\text{MR}}(b, t)$, with $t = 1$, then breakdown in the corresponding single (s, t) -recursion, with $t = 1$ and $s = b$, is only possible for at most a set of \underline{p}_0 of measure zero. In this case, if breakdown occurs for some j , $\underline{w}_{j-1}(\underline{p}_0) = \underline{0}$. We show that it is not possible for $\underline{w}_{j-1}(\underline{p}_0) = 0$ for every \underline{p}_0 .

THEOREM 6.5 Suppose $A \in \mathcal{CG}_{\text{MR}}(b, 1)$. The Hessenberg matrix H_N , with entries, $\{h_{i,j}\}_{i,j=0}^{N-1}$, that results after N steps of a full Gram-Schmidt process, can be written as in (6.1). H_N can be decomposed as

$$H_N = T + U,$$

where T is upper Hessenberg with upper bandwidth b , and U is upper triangular with elements given by (6.2). The vectors, $\underline{w}_j(\underline{p}_0)$ and $\underline{v}_i(\underline{p}_0)$, are of length 1, thus, are just complex scalars.

If there exists $j \geq b + 1$ where breakdown occurs in the single $(b, 1)$ -recursion for every \underline{p}_0 , that is, $\underline{w}_{j-1}(\underline{p}_0) = 0$ for every \underline{p}_0 , then A is B -normal $(b - 1)$ or $d(A) \leq j$.

Proof: First, suppose $b = 0$ and that for some j , $h_{0,j-1} = \langle \underline{w}_{j-1}, \underline{v}_0 \rangle = 0$, for every \underline{p}_0 . Since this must hold for every \underline{p}_0 , it must hold for every \underline{p}_0 such that $d(\underline{p}_0, A) = j$. This means $\{\underline{p}_0, \dots, \underline{p}_{j-1}\}$ span an invariant subspace of A , and

$$AP_j = P_j H_j,$$

where,

$$H_j = \begin{bmatrix} \langle \underline{w}_0, \underline{v}_0 \rangle & \langle \underline{w}_1, \underline{v}_0 \rangle & \cdots & \langle \underline{w}_{j-1}, \underline{v}_0 \rangle \\ 1 & \langle \underline{w}_1, \underline{v}_1 \rangle & & \vdots \\ & \ddots & \ddots & \vdots \\ & & 1 & \langle \underline{w}_{j-1}, \underline{v}_{j-1} \rangle \end{bmatrix}.$$

Since $\underline{w}_{j-1}(\underline{p}_0) = 0$, it follows that the entire j 'th column of H_j is zero. Let $\underline{\epsilon}_j = [0 \cdots 0 \ 1]^T \in \mathcal{C}^j$. It follows that

$$AP_j \underline{\epsilon}_j = P_j H_j \underline{\epsilon}_j = \underline{0}.$$

Since the j 'th column of P_j is \underline{p}_{j-1} ,

$$A \underline{p}_{j-1} = \underline{0},$$

where, $\underline{p}_{j-1} \neq \underline{0}$. This implies A is singular, which is a contradiction.

Suppose $b > 0$, and for some j , $h_{0,j-1} = \langle \underline{w}_{j-1}, \underline{v}_i \rangle = 0$, for every \underline{p}_0 . From the main theorem in [5], if $h_{0,j-1} = 0$, $\forall \underline{p}_0$, then A is B -normal($j-2$), or $d(A) \leq j$.

If $d(A) \leq j$, $\underline{p}_j = \underline{0}$. Breakdown is not a problem because the iteration has already converged. If $j = b+1$, A is B -normal($b-1$).

Suppose that $j > b+1$ and consider the matrix $H_{j+1,j}$ that results after computing \underline{p}_j using (2.8):

$$H_{j+1,j} = \begin{bmatrix} \frac{\langle A \underline{p}_0, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \cdots & \frac{\langle A \underline{p}_{b-1}, \underline{p}_0 \rangle_B}{\langle \underline{p}_0, \underline{p}_0 \rangle_B} & \langle \underline{w}_b, \underline{v}_0 \rangle & \cdots & \langle \underline{w}_{j-1}, \underline{v}_0 \rangle \\ 1 & \frac{\langle A \underline{p}_1, \underline{p}_1 \rangle_B}{\langle \underline{p}_1, \underline{p}_1 \rangle_B} & & \frac{\langle A \underline{p}_b, \underline{p}_1 \rangle_B}{\langle \underline{p}_1, \underline{p}_1 \rangle_B} & \ddots & \vdots \\ & \ddots & \ddots & & \ddots & \langle \underline{w}_{j-1}, \underline{v}_{j-b-1} \rangle \\ & & \ddots & \ddots & \ddots & \frac{\langle A \underline{p}_{j-1}, \underline{p}_{j-b} \rangle_B}{\langle \underline{p}_{j-b}, \underline{p}_{j-b} \rangle_B} \\ & & & \ddots & \ddots & \vdots \\ & & & & 1 & \frac{\langle A \underline{p}_{j-1}, \underline{p}_{j-1} \rangle_B}{\langle \underline{p}_{j-1}, \underline{p}_{j-1} \rangle_B} \\ & & & & & 1 \end{bmatrix}. \quad (6.10)$$

Since $\underline{w}_{j-1}(\underline{p}_0) = 0$, it follows that $h_{i,j-1} = \langle \underline{w}_{j-1}, \underline{v}_i \rangle = 0$, for $i = 0, \dots, j - b - 1$. Notice that for $j > b + 1$, this implies that both

$$\begin{aligned} \langle A\underline{p}_{j-1}, \underline{p}_0 \rangle_B &= \langle \underline{p}_{j-1}, A^\dagger \underline{p}_0 \rangle_B = 0, \quad \text{and} \\ \langle A\underline{p}_{j-1}, \underline{p}_1 \rangle_B &= \langle \underline{p}_{j-1}, A^\dagger \underline{p}_1 \rangle = 0. \end{aligned} \tag{6.11}$$

This must hold for every \underline{p}_0 such that $d(\underline{p}_0, A) = j$. Let

$$\mathcal{V} = \text{sp}\{\underline{p}_0, A\underline{p}_0, \dots, A^{j-1}\underline{p}_0\} = \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{j-1}\}.$$

Since A is B -normal($j - 2$), we know that

$$A^\dagger = p_{j-2}(A),$$

for some polynomial p_{j-2} , and thus,

$$A^\dagger \underline{p}_0 \in \text{sp}\{\underline{p}_0, A\underline{p}_0, \dots, A^{j-2}\underline{p}_0\} = \text{sp}\{\underline{p}_0, \dots, \underline{p}_{j-2}\}. \tag{6.12}$$

Now, $A^\dagger \underline{p}_1 = A^\dagger(A\underline{p}_0 - \sigma_{0,0}\underline{p}_0)$, and by using (6.11) we obtain,

$$\langle \underline{p}_{j-1}, A^\dagger \underline{p}_1 \rangle_B = \langle \underline{p}_{j-1}, A^\dagger A\underline{p}_0 \rangle_B = \langle \underline{p}_{j-1}, AA^\dagger \underline{p}_0 \rangle_B = 0.$$

Notice that

$$AA^\dagger \underline{p}_0 = A(p_{j-2}(A)\underline{p}_0) \in \text{sp}\{\underline{p}_0, A\underline{p}_0, \dots, A^{j-1}\underline{p}_0\} = \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_{j-1}\}.$$

Since $AA^\dagger \underline{p}_0 \sim_B \underline{p}_{j-1}$, it follows that

$$AA^\dagger \underline{p}_0 \in \text{sp}\{\underline{p}_0, A\underline{p}_0, \dots, A^{j-2}\underline{p}_0\}.$$

Using (6.12), we see that the only way this can hold is if

$$A^\dagger \underline{p}_0 \in \text{sp}\{\underline{p}_0, A\underline{p}_0, \dots, A^{j-3}\underline{p}_0\}, \quad \forall \underline{p}_0 \text{ such that } d(\underline{p}_0, A) = j.$$

From Lemma 2 in [5], it follows that $A|_{\mathcal{V}}$ is B -normal($j - 3$), where $A|_{\mathcal{V}}$ denotes the restriction of A to \mathcal{V} . Since this holds for every \underline{p}_0 with $d(\underline{p}_0, A) = j$, it follows from the main proof in [5], that A is B -normal($j - 3$), which implies

$$h_{0,j-2} = \langle \underline{w}_{j-2}, \underline{v}_0 \rangle = 0, \quad \forall \underline{p}_0.$$

This means that breakdown must occur at the previous step, and $\underline{w}_{j-2}(\underline{p}_0) = 0, \forall \underline{p}_0$. If $j = b + 2$, A is B -normal($b - 1$), and we are done. If $j > b + 2$, the argument can be repeated showing that A is really B -normal($j - 4$). This process can be continued until we obtain that A is B -normal($b - 1$). \square

From Theorem 6.5, we see that if $A \in \mathcal{CG}_{\text{MR}}(b, t)$, with $t = 1$, then either A is B -normal($b - 1$) and a single (s, t) -recursion with $(s, t) = (b - 1, 0)$ yields a B -orthogonal basis, or a single (s, t) -recursion, with $(s, t) = (b, 1)$, will yield this basis for all except possibly a measure zero set of \underline{p}_0 .

DEFINITION 6.2 $A \in \mathcal{CG}_{\text{SR}}(s, t)$ if for almost every \underline{p}_0 and every $0 \leq j \leq d(\underline{p}_0, A) - 2$, there exists coefficients $\{\beta_{k,j}\}_{k=j-t}^{j-1}$ such that

$$\sigma_{i,j} = \frac{\langle \underline{p}_j + \sum_{k=j-t}^{j-1} \beta_{k,j} \underline{p}_k, A^\dagger \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B} = 0, \quad \text{for } i = 0, \dots, j - s - 1.$$

In other words, $A \in \mathcal{CG}_{\text{SR}}(s, t)$, if for almost every \underline{p}_0 , an (s, t) -recursion (6.5) can be used to construct a B -orthogonal basis for $\mathcal{K}_d(\underline{r}_0, A)$.

COROLLARY 6.6 If $A \in \mathcal{CG}_{\text{MR}}(b, t)$, with $t \leq 1$, then either $d(A) \leq b + 1$, $A \in \mathcal{CG}_{\text{SR}}(s, t)$ with $(s, t) = (b, 1)$, or $A \in \mathcal{CG}_{\text{SR}}(s, t)$ with $(s, t) = (b - 1, 0)$.

Proof: If $t = 0$, the discussion before Theorem 6.5 shows that $d(A) \leq b + 1$, or A is B -normal($b - 1$). If A is B -normal($b - 1$), then $A \in \mathcal{CG}_{\text{SR}}(b - 1, 0)$, or $d(A) \leq b + 1$.

If $t = 1$, Theorem 6.5 says that if $\underline{w}_{j-1}(\underline{p}_0) = 0$, for every \underline{p}_0 , then $d(A) \leq b + 1$, or A is B -normal($b - 1$), which implies $A \in \mathcal{CG}_{\text{SR}}(s, t)$, with $(s, t) = (b - 1, 0)$. Otherwise, $\underline{w}_{j-1}(\underline{p}_0) = 0$ on at most a set of \underline{p}_0 of measure zero, thus, $A \in \mathcal{CG}_{\text{SR}}(s, t)$, with $(s, t) = (b, 1)$. \square

Next, we review some basic concepts from linear algebra. These concepts will be used to develop a tool that will be used in the analysis of the class $\mathcal{CG}_{\text{SR}}(s, t)$.

Given a set of s vectors of length N , $\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_s\}$, recall that the wedge product ([21], pp. 553-560)

$$\mathcal{F}(\underline{v}_1, \underline{v}_2, \dots, \underline{v}_s) = \underline{v}_1 \wedge \underline{v}_2 \wedge \dots \wedge \underline{v}_s,$$

(where \wedge denotes the wedge product), is a multilinear function from $\mathcal{C}^{N \times s}$ to $\mathcal{C}^{\binom{N}{s}}$. \mathcal{F} maps the set of vectors $\{\underline{v}_1, \dots, \underline{v}_s\}$ written as the $N \times s$ matrix, $V_{N,s} = [\underline{v}_1 \ \underline{v}_2 \ \dots \ \underline{v}_s]$, onto a single vector \underline{v} of length $\binom{N}{s}$. Each component of \underline{v} is the result of taking the determinant of an $s \times s$ matrix formed by choosing s of the N rows of $V_{N,s}$. Since there are $\binom{N}{s}$ possible ways of doing this, \underline{v} has length $\binom{N}{s}$. If $s = N$, $\mathcal{F}(\underline{v}_1, \dots, \underline{v}_s)$ is just the determinant of the $N \times N$ matrix V_N . Since

$$\mathcal{F}(\underline{v}_1, \dots, \underline{v}_s) \equiv \underline{0} \iff \{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_s\} \text{ are linearly dependent,}$$

the wedge product can be used to determine the linear independence (dependence) of a set of vectors.

A list of properties of determinants that will be used throughout the main proofs in this chapter is given in Appendix A. In the following discussion, we will denote

$$|A|, \quad \text{as the determinant of the matrix } A.$$

A less well known result is stated below.

Denote $a_{i,j}(t)$ to be the entries of a $p \times p$ matrix $A(t)$. Then, $(\partial/\partial t)|A(t)|$ is the sum of the p determinants,

$$\begin{vmatrix} a'_{1,1}(t) & a_{1,2}(t) & \cdots & a_{1,p}(t) \\ \vdots & \vdots & & \vdots \\ a'_{p,1}(t) & a_{p,2}(t) & \cdots & a_{p,p}(t) \end{vmatrix} + \cdots + \begin{vmatrix} a_{1,1}(t) & \cdots & a_{1,p-1}(t) & a'_{1,p}(t) \\ \vdots & & \vdots & \vdots \\ a_{p,1}(t) & \cdots & a_{p,p-1}(t) & a'_{p,p}(t) \end{vmatrix}. \quad (6.13)$$

In the following Lemma, a useful tool for analyzing the class $\mathcal{CG}_{\text{SR}}(s, t)$ is derived.

LEMMA 6.7 If $A \in \mathcal{CG}_{\text{SR}}(s, t)$, then for every $\underline{p}_0 \in \mathcal{C}^N$ such that $d(\underline{p}_0) = s + t + 2$, the wedge product,

$$\mathcal{F}(\underline{p}_0, \tilde{A}) = \underline{p}_0 \wedge \tilde{A}\underline{p}_0 \wedge \cdots \wedge \tilde{A}^s \underline{p}_0 \wedge \tilde{A}^\dagger \underline{p}_0 \wedge \tilde{A}^\dagger \tilde{A}\underline{p}_0 \wedge \cdots \wedge \tilde{A}^\dagger \tilde{A}^t \underline{p}_0 = \underline{0}, \quad (6.14)$$

where \tilde{A} denotes the restriction of A to the subspace,

$$\mathcal{V}_{s+t+2} = \text{sp}\{\underline{p}_0, A\underline{p}_0, \dots, A^{s+t+1}\}. \quad (6.15)$$

Proof: Suppose $A \in \mathcal{CG}_{\text{SR}}(s, t)$. From Definition 6.2 it follows that for almost every \underline{p}_0 , an (s, t) -recursion yields a B -orthogonal basis for $\mathcal{K}_d(\underline{p}_0, A)$. Since almost every vector in \mathcal{C}^N has degree $d = d(A)$ (see [19], pp. 62), and $d(A) > s + t + 2$, it follows that for almost every \underline{p}_0 with $d(\underline{p}_0) > s + t + 2$, an (s, t) -recursion (6.5) yields a B -orthogonal basis. For any \underline{p}_0 that satisfies this condition, the (s, t) -recursion yields $\underline{p}_{s+t+2} \neq \underline{0}$. By substituting $j + 1 = s + t + 2$ in (6.5) we obtain,

$$\underline{p}_{s+t+2} = A\underline{p}_{s+t+1} + \sum_{k=s+1}^{s+t} \beta_{k,(s+t+1)} A\underline{p}_k - \sum_{i=t+1}^{s+t+1} \sigma_{i,(s+t+1)} \underline{p}_i. \quad (6.16)$$

There exists coefficients, $\{\beta_{k,(s+t+1)}\}_{k=s+1}^{s+t}$, such that

$$\sigma_{i,(s+t+1)} = \frac{\langle A\underline{p}_{s+t+1} + \sum_{k=s+1}^{s+t} \beta_{k,(s+t+1)} A\underline{p}_k, \underline{p}_i \rangle_B}{\langle \underline{p}_i, \underline{p}_i \rangle_B} = 0, \quad \text{for } i < t + 1.$$

Equivalently, we denote

$$\mathcal{W}_i(\underline{p}_0) = \langle \underline{p}_{s+t+1} + \sum_{k=s+1}^{s+t} \beta_{k,(s+t+1)} \underline{p}_k, A^\dagger \underline{p}_i \rangle_B = 0, \quad \text{for } i < t + 1. \quad (6.17)$$

For $i = 0, \dots, t$, $\mathcal{W}_i(\underline{p}_0)$ can be written as the system of equations:

$$\begin{bmatrix} \langle \underline{p}_{s+t+1}, A^\dagger \underline{p}_0 \rangle_B & \cdots & \langle \underline{p}_{s+1}, A^\dagger \underline{p}_0 \rangle_B \\ \vdots & & \vdots \\ \langle \underline{p}_{s+t+1}, A^\dagger \underline{p}_t \rangle_B & \cdots & \langle \underline{p}_{s+1}, A^\dagger \underline{p}_t \rangle_B \end{bmatrix} \begin{pmatrix} 1 \\ \beta_{s+t,s+t+1} \\ \cdots \\ \beta_{s+1,s+t+1} \end{pmatrix} = \underline{0}.$$

Notice that this is a $(t + 1) \times (t + 1)$ system, with a nontrivial nullspace.

Summarizing the above, we see that if $A \in \mathcal{CG}_{\text{SR}}(s, t)$, then for almost every \underline{p}_0 with $d(\underline{p}_0) > s + t + 2$, $\mathcal{W}_i(\underline{p}_0) = 0$, for $i = 0, \dots, t$. In [5] it was proven that for every j , \underline{p}_j is a continuous function of \underline{p}_0 . Since the inner product is also continuous, it follows that for $i = 0, \dots, t$, $\mathcal{W}_i(\underline{p}_0)$ is a continuous function of \underline{p}_0 . The set for which $d(\underline{p}_0) \leq s + t + 2$ is closed and of smaller dimension than the set for which $d(\underline{p}_0) > s + t + 2$. Therefore, the set of \underline{p}_0 such that $d(\underline{p}_0) \leq s + t + 2$ is a set of measure zero in \mathcal{C}^N . For $i = 0, \dots, t$, we know that \mathcal{W}_i must be zero on all but a set of initial vectors, \underline{p}_0 , of measure zero. Since \mathcal{W}_i is a continuous function of \underline{p}_0 , \mathcal{W}_i must be zero everywhere. Therefore, $\mathcal{W}_i(\underline{p}_0) = 0$, $\forall \underline{p}_0$ such that $d(\underline{p}_0) \leq s + t + 2$.

Suppose $d(\underline{p}_0) = s + t + 2$, and let

$$\mathcal{V}_{s+t+2} = \text{sp}\{\underline{p}_0, A\underline{p}_0, \dots, A^{s+t+1}\underline{p}_0\} = \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_s, \underline{p}_{s+1}, \dots, \underline{p}_{s+t+1}\}.$$

\mathcal{V}_{s+t+2} is an A invariant subspace of \mathcal{C}^N . Denote

$$V = \begin{bmatrix} | & | & & | \\ \underline{p}_0 & \underline{p}_1 & \cdots & \underline{p}_{s+t+1} \\ | & | & & | \end{bmatrix},$$

as the matrix whose columns are the B -orthogonal basis vectors for \mathcal{V}_{s+t+2} , and let

$$U = V(V^*BV)^{-1}V^*B,$$

be the B -orthogonal projector onto \mathcal{V}_{s+t+2} ([15], pp. 8-10). It follows that for every $\underline{v} \in \mathcal{C}^N$, $U\underline{v} \in \mathcal{V}_{s+t+2}$, and furthermore, if $\underline{v} \in \mathcal{V}_{s+t+2}$, then $U\underline{v} = \underline{v}$.

Let \tilde{A} be the restriction of A to \mathcal{V}_{s+t+2} . We can write,

$$\tilde{A} = AU.$$

Notice that

$$\tilde{A}^\dagger = U^\dagger A^\dagger, \quad \text{and} \quad U^\dagger = B^{-1}U^*B = B^{-1}[V(V^*BV)^{-1}V^*B]^*B = U,$$

so it follows that

$$\tilde{A}^\dagger = UA^\dagger.$$

Therefore, \mathcal{V}_{s+t+2} is an invariant subspace of both \tilde{A} and \tilde{A}^\dagger . For every \underline{p}_0 such that $d(\underline{p}_0) = s + t + 2$, and for $i = 0, \dots, t$,

$$\begin{aligned} \mathcal{W}_i(\underline{p}_0) &= \langle A\underline{p}_{s+t+1} + \sum_{k=s+1}^{s+t} \beta_{k,(s+t+1)} A\underline{p}_k, \underline{p}_i \rangle_B \\ &= \langle \tilde{A}\underline{p}_{s+t+1} + \sum_{k=s+1}^{s+t} \beta_{k,(s+t+1)} \tilde{A}\underline{p}_k, \underline{p}_i \rangle_B \\ &= \langle \underline{p}_{s+t+1} + \sum_{k=s+1}^{s+t} \beta_{k,(s+t+1)} \underline{p}_k, \tilde{A}^\dagger \underline{p}_i \rangle_B = 0. \end{aligned}$$

This means that $\text{sp}\{\tilde{A}^\dagger \underline{p}_0, \tilde{A}^\dagger \underline{p}_1, \dots, \tilde{A}^\dagger \underline{p}_t\}$ is B -orthogonal to the vector,

$$\hat{\underline{p}}_0 = \underline{p}_{s+t+1} + \sum_{k=s+1}^{s+t} \beta_{k,(s+t+1)} \underline{p}_k \in \text{sp}\{\underline{p}_{s+1}, \underline{p}_{s+2}, \dots, \underline{p}_{s+t+1}\}.$$

Furthermore, since \mathcal{V}_{s+t+2} is an invariant subspace of \tilde{A}^\dagger ,

$$\text{sp}\{\tilde{A}^\dagger \underline{p}_0, \tilde{A}^\dagger \underline{p}_1, \dots, \tilde{A}^\dagger \underline{p}_t\} \subset \mathcal{V}_{s+t+2}.$$

Notice that $\hat{\underline{p}}_0$ is also B -orthogonal to $\text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_s\}$. Starting with $\hat{\underline{p}}_0$, take the set, $\{\hat{\underline{p}}_0, \underline{p}_{s+1}, \dots, \underline{p}_{s+t+1}\}$, of $t + 2$ vectors and use a Gram-Schmidt process to construct a set of $t+1$ vectors that form a B -orthogonal basis for $\text{sp}\{\hat{\underline{p}}_0, \underline{p}_{s+1}, \dots, \underline{p}_{s+t+1}\}$.

Call this set

$$\hat{\mathcal{V}}_{t+1} = \text{sp}\{\hat{\underline{p}}_0, \hat{\underline{p}}_1, \dots, \hat{\underline{p}}_t\}.$$

Together, $\text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_s\}$ and $\text{sp}\{\hat{\underline{p}}_0, \hat{\underline{p}}_1, \dots, \hat{\underline{p}}_t\}$ form a B -orthogonal basis for

$$\mathcal{V}_{s+t+2} = \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_s, \hat{\underline{p}}_0, \hat{\underline{p}}_1, \dots, \hat{\underline{p}}_t\}.$$

Since $\{\tilde{A}^\dagger \underline{p}_0, \tilde{A}^\dagger \underline{p}_1, \dots, \tilde{A}^\dagger \underline{p}_t\}$ are all B -orthogonal to $\hat{\underline{p}}_0$, we must have both

$$\{\tilde{A}^\dagger \underline{p}_0, \tilde{A}^\dagger \underline{p}_1, \dots, \tilde{A}^\dagger \underline{p}_t\} \subset \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_s, \hat{\underline{p}}_1, \dots, \hat{\underline{p}}_t\}$$

and

$$\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_s\} \subset \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_s, \hat{\underline{p}}_1, \dots, \hat{\underline{p}}_t\}.$$

This means there are $s+t+2$ vectors, all included in the span of an $s+t+1$ dimensional subspace of \mathcal{V}_{s+t+2} . Therefore, the vectors, $\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_s, \tilde{A}^\dagger \underline{p}_0, \tilde{A}^\dagger \underline{p}_1, \dots, \tilde{A}^\dagger \underline{p}_t\}$ must be linearly dependent. Since

$$\begin{aligned} & \text{sp}\{\underline{p}_0, \underline{p}_1, \dots, \underline{p}_s, \tilde{A}^\dagger \underline{p}_0, \tilde{A}^\dagger \underline{p}_1, \dots, \tilde{A}^\dagger \underline{p}_t\} \\ &= \text{sp}\{\underline{p}_0, \tilde{A} \underline{p}_0, \dots, \tilde{A}^s \underline{p}_0, \tilde{A}^\dagger \underline{p}_0, \tilde{A}^\dagger \tilde{A} \underline{p}_0, \dots, \tilde{A}^\dagger \tilde{A}^t \underline{p}_0\}, \end{aligned}$$

it follows that for every \underline{p}_0 such that $d(\underline{p}_0) = s+t+2$, the wedge product

$$\mathcal{F}(\underline{p}_0, \tilde{A}) = \underline{p}_0 \wedge \tilde{A} \underline{p}_0 \wedge \dots \wedge \tilde{A}^s \underline{p}_0 \wedge \tilde{A}^\dagger \underline{p}_0 \wedge \tilde{A}^\dagger \tilde{A} \underline{p}_0 \wedge \dots \wedge \tilde{A}^\dagger \tilde{A}^t \underline{p}_0 = \underline{0}.$$

□

Let \mathcal{T} be a linear transformation from a q -dimensional vector space, \mathcal{X} , to a p -dimensional vector space, \mathcal{Y} . Recall that once we choose ordered bases, $\{\underline{x}_i\}_{i=1}^q$ for \mathcal{X} , and $\{\underline{y}_i\}_{i=1}^p$ for \mathcal{Y} , that isomorphisms are induced from \mathcal{X} to \mathcal{C}^q , and from \mathcal{Y} to \mathcal{C}^p , and a matrix representation of the linear transformation, \mathcal{T} , given by a $p \times q$ matrix T . ([22], pp. 201, pp. 257-262).

In this setting, we can think of \tilde{A} as defining a linear transformation from an $(s+t+2)$ -dimensional subspace, \mathcal{V}_{s+t+2} of \mathcal{C}^N , to \mathcal{V}_{s+t+2} . Since B is a Hermitian positive definite matrix,

$$\langle \cdot, \cdot \rangle = \langle B \cdot, \cdot \rangle,$$

defines an inner product on \mathcal{V}_{s+t+2} . It follows from Schur's theorem ([22], pp 326-328), that there is an orthonormal basis for \mathcal{V}_{s+t+2} , with respect to $\langle B \cdot, \cdot \rangle$, for which the matrix representation of \tilde{A} is an $(s+t+2) \times (s+t+2)$ upper triangular matrix, \tilde{R} . This means there exists a bijective map \tilde{Q} from \mathcal{C}^{s+t+2} to \mathcal{V}_{s+t+2} , and thus, \tilde{Q}^{-1} , for which

$$\tilde{A} = \tilde{Q} \tilde{R} \tilde{Q}^{-1}, \quad \text{and} \quad \tilde{Q}^* B \tilde{Q} = I \in \mathcal{C}^{(s+t+2) \times (s+t+2)}. \quad (6.18)$$

Recall that $\tilde{A}^\dagger = B^{-1}\tilde{A}^*B$. From (6.18), we see that $\tilde{Q}^*B = \tilde{Q}^{-1}$. Using this information upon substitution of $\tilde{A} = \tilde{Q}\tilde{R}\tilde{Q}^{-1}$ into (6.14), we obtain

$$\mathcal{F}(\underline{p}_0, \tilde{A}) = \underline{p}_0 \wedge \tilde{Q}\tilde{R}\tilde{Q}^{-1}\underline{p}_0 \wedge \cdots \wedge \tilde{Q}\tilde{R}^s\tilde{Q}^{-1}\underline{p}_0 \wedge \tilde{Q}\tilde{R}^*\tilde{Q}^{-1}\underline{p}_0 \wedge \cdots \wedge \tilde{Q}\tilde{R}^*\tilde{R}^t\tilde{Q}^{-1}\underline{p}_0 = \underline{0},$$

for every $\underline{p}_0 \in \mathcal{C}^N$ such that $d(\underline{p}_0, A) = s + t + 2$. Since every $\underline{w} \in \mathcal{C}^{s+t+2}$ can be written as,

$$\underline{w} = \tilde{Q}^{-1}\underline{p}_0,$$

for some \underline{p}_0 , we rewrite the above wedge product as,

$$\mathcal{F}(\underline{p}_0, \tilde{A}) = \tilde{Q}\underline{w} \wedge \tilde{Q}\tilde{R}\underline{w} \wedge \cdots \wedge \tilde{Q}\tilde{R}^s\underline{w} \wedge \tilde{Q}\tilde{R}^*\underline{w} \wedge \cdots \wedge \tilde{Q}\tilde{R}^*\tilde{R}^t\underline{w} = \underline{0},$$

for every $\underline{w} \in \mathcal{C}^{s+t+2}$. Since \tilde{Q} has linearly independent columns, it follows that $\mathcal{F}(\underline{p}_0, \tilde{A}) = \underline{0}$, for every $\underline{p}_0 \in \mathcal{C}^N$ such that $d(\underline{p}_0, A) = s + t + 2$, if and only if,

$$\mathcal{F}(\underline{w}, \tilde{R}) = \underline{w} \wedge \tilde{R}\underline{w} \wedge \cdots \wedge \tilde{R}^s\underline{w} \wedge \tilde{R}^*\underline{w} \wedge \tilde{R}^*\tilde{R}\underline{w} \wedge \cdots \wedge \tilde{R}^*\tilde{R}^t\underline{w} = \underline{0},$$

for every $\underline{w} \in \mathcal{C}^{s+t+2}$. This observation allows us to assume A is an $(s + t + 2) \times (s + t + 2)$ upper triangular matrix.

6.3 Proof Of Necessary Conditions

We are now ready to prove the main theoretical result regarding necessary conditions for $A \in \mathcal{CG}_{\text{MR}}(b, t)$. This is a difficult problem, and in this chapter the analysis will be limited to a subset of the class $\mathcal{CG}_{\text{MR}}(b, t)$. In particular, we will consider the class $\mathcal{CG}_{\text{MR}}(b, t)$, with $(b, t) \leq (1, 1)$. We begin by proving a result concerning the class $\mathcal{CG}_{\text{SR}}(s, t)$, with $(s, t) \leq (1, 1)$. In Corollary 6.9, the corresponding result for the class $\mathcal{CG}_{\text{MR}}(b, t)$, with $(b, t) \leq (1, 1)$, is given.

THEOREM 6.8 $A \in \mathcal{CG}_{\text{SR}}(s, t)$, with $(s, t) \leq (1, 1)$, if and only if

- (1) $d(A) \leq 3$, or

(2) A is B -normal(s, t), with $(s, t) \leq (1, 1)$.

Proof: Sufficient conditions have been partially established in Chapter 4. If $d(A) \leq 3$, $d(\underline{p}_0, A) \leq 3$ for every \underline{p}_0 , and a single $(1, 1)$ -recursion can always be used to construct a B -orthogonal basis for $\mathcal{K}_d(r_0, A) = \text{sp}\{\underline{p}_0, \underline{p}_1, \underline{p}_2\}$. From Section 4.3, we know that if A is B -normal(s, t), a single (s, t) -recursion can be used to construct a B -orthogonal basis if the system given in (4.13) is consistent at each step. When $(s, t) \leq (1, 1)$, it follows from the proof of Theorem 6.5 that the matrix in (4.13) can only be singular for a set of \underline{p}_0 of measure zero. Therefore, by Definition 6.2, $A \in \mathcal{CG}_{\text{SR}}(s, t)$, with $(s, t) \leq (1, 1)$.

Suppose $A \in \mathcal{CG}_{\text{SR}}(1, 1)$. Choose any $\underline{p}_0 \in \mathcal{C}^N$ with $d(\underline{p}_0, A) = 4$, and denote \tilde{A} as the restriction of A to \mathcal{V}_4 , where

$$\mathcal{V}_4 = \text{sp}\{\underline{p}_0, A\underline{p}_0, A^2\underline{p}_0, A^3\underline{p}_0\}.$$

From Lemma 6.7, it follows that for every $\underline{p}_0 \in \mathcal{C}^N$, such that $d(\underline{p}_0, A) = 4$,

$$\mathcal{F}(\underline{p}_0, \tilde{A}) = \underline{p}_0 \wedge \tilde{A}\underline{p}_0 \wedge \tilde{A}^\dagger \underline{p}_0 \wedge \tilde{A}^\dagger \tilde{A}\underline{p}_0 = 0. \quad (6.19)$$

The discussion following Lemma 6.7 shows the wedge product (6.19) holds if and only if

$$\mathcal{F}(\underline{w}, \tilde{R}) = \underline{w} \wedge \tilde{R}\underline{w} \wedge \tilde{R}^* \underline{w} \wedge \tilde{R}^* \tilde{R}\underline{w} = 0, \quad (6.20)$$

for every $\underline{w} \in \mathcal{C}^4$, where \tilde{R} is the upper triangular, matrix representation of \tilde{A} that results from applying Schur's Theorem.

Notice that \tilde{R} is a 4×4 upper triangular matrix. Since \tilde{R} is a square matrix, the wedge product reduces to a single determinant,

$$|\underline{w}, \tilde{R}\underline{w}, \tilde{R}^* \underline{w}, \tilde{R}^* \tilde{R}\underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4. \quad (6.21)$$

In order to obtain a more useful form for our analysis, we will perform a series of operations on (6.21) using the properties of determinants listed in Appendix

A.

The determinant in (6.21) does not change if we add a multiple of one column to another column. Let λ_j be an eigenvalue of \tilde{R} . Multiplying the third column in (6.21) by $(-\lambda_j)$ and adding it to the fourth column yields,

$$|\underline{\tilde{w}}, \tilde{R}\underline{\tilde{w}}, \tilde{R}^*\underline{\tilde{w}}, \tilde{R}^*(\tilde{R} - \lambda_j I)\underline{\tilde{w}}| = 0, \quad \forall \underline{\tilde{w}} \in \mathcal{C}^4.$$

Similarly, multiplying the first column in the above determinant by $(-\lambda_j)$ and adding it to the second column, then multiplying the first column by $(-\bar{\lambda}_j)$ and adding it to the third column produces,

$$|\underline{\tilde{w}}, (\tilde{R} - \lambda_j I)\underline{\tilde{w}}, (\tilde{R}^* - \bar{\lambda}_j I)\underline{\tilde{w}}, \tilde{R}^*(\tilde{R} - \lambda_j I)\underline{\tilde{w}}| = 0, \quad \forall \underline{\tilde{w}} \in \mathcal{C}^4.$$

Finally, multiplying the second column by $(-\bar{\lambda}_j)$ and adding it to the fourth column yields,

$$|\underline{\tilde{w}}, (\tilde{R} - \lambda_j I)\underline{\tilde{w}}, (\tilde{R}^* - \bar{\lambda}_j I)\underline{\tilde{w}}, (\tilde{R}^* - \bar{\lambda}_j I)(\tilde{R} - \lambda_j I)\underline{\tilde{w}}| = 0, \quad \forall \underline{\tilde{w}} \in \mathcal{C}^4.$$

To simplify the notation, we denote

$$G_j = (\tilde{R} - \lambda_j I), \quad \text{and} \quad G_j^* = (\tilde{R}^* - \bar{\lambda}_j I),$$

and rewrite the above determinant as,

$$|\underline{\tilde{w}}, G_j \underline{\tilde{w}}, G_j^* \underline{\tilde{w}}, G_j^* G_j \underline{\tilde{w}}| = 0, \quad \forall \underline{\tilde{w}} \in \mathcal{C}^4. \quad (6.22)$$

Every $\underline{\tilde{w}} \in \mathcal{C}^4$ can be written as $\underline{\tilde{w}} = \underline{u} + \alpha \underline{v}$ for some $\underline{u}, \underline{v} \in \mathcal{C}^4$ and for some $\alpha \in \mathcal{C}$. Denote

$$\mathcal{G}(\alpha) = |\underline{u} + \alpha \underline{v}, G_j(\underline{u} + \alpha \underline{v}), G_j^*(\underline{u} + \alpha \underline{v}), G_j^* G_j(\underline{u} + \alpha \underline{v})|, \quad (6.23)$$

and note that $\mathcal{G}(\alpha) = 0$ for every $\underline{u}, \underline{v} \in \mathcal{C}^4$, and every $\alpha \in \mathcal{C}$. Since the determinant, $\mathcal{G}(\alpha)$, can be written as a polynomial in α , and this polynomial is zero everywhere, it

follows that all derivatives of $\mathcal{G}(\alpha)$ must also be zero. Differentiating \mathcal{G} with respect to α and using (6.13) yields:

$$\begin{aligned}\mathcal{G}'(\alpha) &= |\underline{v}, G_j(\underline{u} + \alpha\underline{v}), G_j^*(\underline{u} + \alpha\underline{v}), G_j^*G_j(\underline{u} + \alpha\underline{v})| \\ &+ |\underline{u} + \alpha\underline{v}, G_j\underline{v}, G_j^*(\underline{u} + \alpha\underline{v}), G_j^*G_j(\underline{u} + \alpha\underline{v})| \\ &+ |\underline{u} + \alpha\underline{v}, G_j(\underline{u} + \alpha\underline{v}), G_j^*\underline{v}, G_j^*G_j(\underline{u} + \alpha\underline{v})| \\ &+ |\underline{u} + \alpha\underline{v}, G_j(\underline{u} + \alpha\underline{v}), G_j^*(\underline{u} + \alpha\underline{v}), G_j^*G_j\underline{v}| = 0,\end{aligned}$$

for every $\underline{u}, \underline{v} \in \mathcal{C}^4$, and for every $\alpha \in \mathcal{C}$. Differentiating again with respect to α , and then setting $\alpha = 0$, produces:

$$\begin{aligned}\mathcal{G}''(\alpha)|_{\alpha=0} &= 2|\underline{v}, G_j\underline{v}, G_j^*\underline{u}, G_j^*G_j\underline{u}| + 2|\underline{v}, G_j\underline{u}, G_j^*\underline{v}, G_j^*G_j\underline{u}| \\ &+ 2|\underline{v}, G_j\underline{u}, G_j^*\underline{u}, G_j^*G_j\underline{v}| + 2|\underline{u}, G_j\underline{v}, G_j^*\underline{v}, G_j^*G_j\underline{u}| \\ &+ 2|\underline{u}, G_j\underline{v}, G_j^*\underline{u}, G_j^*G_j\underline{v}| + 2|\underline{u}, G_j\underline{u}, G_j^*\underline{v}, G_j^*G_j\underline{v}| = 0,\end{aligned}$$

for every $\underline{u}, \underline{v} \in \mathcal{C}^4$. For any $\underline{u} \in \mathcal{C}^4$, these twelve determinants must add up to zero for every $\underline{v} \in \mathcal{C}^4$. Choose $\underline{u} = \underline{p}_j$, where \underline{p}_j is an eigenvector of \tilde{R} with eigenvalue λ_j , that is,

$$\tilde{R}\underline{p}_j = \lambda_j\underline{p}_j, \quad \text{or} \quad G_j\underline{p}_j = \underline{0}.$$

By making this substitution for \underline{u} into the above, and deleting those determinants that are zero, we obtain

$$2|\underline{p}_j, G_j\underline{v}, G_j^*\underline{p}_j, G_j^*G_j\underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4.$$

Notice that we can switch the order of the second and third columns, and then divide both sides by -2 to produce

$$|\underline{p}_j, G_j^*\underline{p}_j, G_j\underline{v}, G_j^*G_j\underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4. \quad (6.24)$$

Next, by multiplying the first column by $\bar{\lambda}_j$ and adding it to the second column, then multiplying the third column by $\bar{\lambda}_j$ and adding it to the fourth column yields,

$$|\underline{p}_j, \tilde{R}^*\underline{p}_j, (\tilde{R} - \lambda_j I)\underline{v}, \tilde{R}^*(\tilde{R} - \lambda_j I)\underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4. \quad (6.25)$$

The forms given in (6.24) and (6.25) will be useful in the analysis that follows.

For clarity of this presentation, we will divide the remainder of the proof into 2 parts. In Part A, we assume that A is diagonalizable. Part B extends the proof to include general matrices, by showing if $A \in \mathcal{CG}_{\text{SR}}(1, 1)$, and $d(A) > 3$, then A cannot have any nonlinear elementary divisors.

Part A: Suppose that A is diagonalizable.

Since A is diagonalizable, it follows that all the eigenvalues of \tilde{A} and thus of \tilde{R} are simple. Furthermore, since $d(\underline{p}_0, A) = 4$, \tilde{A} has 4 distinct eigenvalues. We begin by showing \tilde{R} is I -normal. To do this, we must show that each eigenvector of \tilde{R} , which we denote as \underline{p}_j with eigenvalue λ_j , is also an eigenvector of \tilde{R}^* with eigenvalue $\bar{\lambda}_j$. Define $\underline{q}_1, \dots, \underline{q}_4$, to be the eigenvectors of \tilde{R}^* such that

$$\langle \underline{q}_i, \underline{p}_j \rangle = 0, \quad \text{if } i \neq j.$$

Let \underline{p}_1 be an eigenvector of \tilde{R} with eigenvalue λ_1 . Recall that (6.25) must hold for $\underline{p}_j = \underline{p}_1$, and $\lambda_j = \lambda_1$, so that

$$|\underline{p}_1, \tilde{R}^* \underline{p}_1, (\tilde{R} - \lambda_1 I) \underline{v}, \tilde{R}^* (\tilde{R} - \lambda_1 I) \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4. \quad (6.26)$$

Denote

$\mathcal{R}(T)$ as the range of T , and

$\mathcal{N}(T)$ as the nullspace of T .

Since \underline{p}_1 is a simple eigenvector, $\mathcal{R}(\tilde{R} - \lambda_1 I)$ is the orthogonal compliment of $\mathcal{N}(\tilde{R}^* - \bar{\lambda}_1 I) = \text{sp}\{\underline{q}_1\}$, and thus, $\mathcal{R}(\tilde{R} - \lambda_1 I) = \text{sp}\{\underline{p}_2, \underline{p}_3, \underline{p}_4\}$. Every $\underline{w} \in \mathcal{C}^4$ can be written as,

$$\underline{w} = (\tilde{R} - \lambda_1 I) \underline{v} + \gamma \underline{p}_1,$$

for some $\underline{y} \in \mathcal{C}^4$ and some $\gamma \in \mathcal{C}$. In (6.26), we add γ times the first column to the third column, and then add γ times the second column to the fourth column to obtain

$$|\underline{p}_1, \tilde{R}^* \underline{p}_1, \underline{w}, \tilde{R}^* \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4. \quad (6.27)$$

Further, any $\underline{w} \in \mathcal{C}^4$ can be written as $\underline{w} = \underline{y} + \alpha \underline{z}$, for some $\underline{y}, \underline{z} \in \mathcal{C}^4$, and some $\alpha \in \mathcal{C}$. Denote

$$\mathcal{G}(\alpha) = |\underline{p}_1, \tilde{R}^* \underline{p}_1, \underline{y} + \alpha \underline{z}, \tilde{R}^* (\underline{y} + \alpha \underline{z})|,$$

and note that $\mathcal{G}(\alpha) = 0$, for every $\underline{y}, \underline{z} \in \mathcal{C}^4$, and every $\alpha \in \mathcal{C}$. Since $\mathcal{G}(\alpha)$ can be written as a polynomial in α , and this polynomial is zero everywhere, it follows that $\mathcal{G}'(\alpha)$ must also be zero. Taking the first derivative with respect to α , then setting $\alpha = 0$, yields

$$\mathcal{G}'(\alpha)|_{\alpha=0} = |\underline{p}_1, \tilde{R}^* \underline{p}_1, \underline{y}, \tilde{R}^* \underline{z}| + |\underline{p}_1, \tilde{R}^* \underline{p}_1, \underline{z}, \tilde{R}^* \underline{y}| = 0, \quad \forall \underline{y}, \underline{z} \in \mathcal{C}^4. \quad (6.28)$$

Choosing $\underline{y} = \underline{q}_2$, where $R^* \underline{q}_2 = \bar{\lambda}_2 \underline{q}_2$, and substituting into (6.28) produces

$$|\underline{p}_1, \tilde{R}^* \underline{p}_1, \underline{q}_2, \tilde{R}^* \underline{z}| + |\underline{p}_1, \tilde{R}^* \underline{p}_1, \underline{z}, \bar{\lambda}_2 \underline{q}_2| = 0, \quad \forall \underline{z} \in \mathcal{C}^4.$$

The properties of determinants can be used to rearrange and combine this into

$$|\underline{p}_1, \tilde{R}^* \underline{p}_1, \underline{q}_2, (\tilde{R}^* - \bar{\lambda}_2 I) \underline{z}| = 0, \quad \forall \underline{z} \in \mathcal{C}^4. \quad (6.29)$$

Since \underline{q}_2 is a simple eigenvector of \tilde{R}^* with eigenvalue $\bar{\lambda}_2$, $\mathcal{R}(\tilde{R}^* - \bar{\lambda}_2 I)$ is the orthogonal complement of $\mathcal{N}(\tilde{R} - \lambda_2 I) = \text{sp}\{\underline{p}_2\}$. Thus, $\mathcal{R}(\tilde{R}^* - \bar{\lambda}_2 I) = \text{sp}\{\underline{q}_1, \underline{q}_3, \underline{q}_4\}$. By multiplying the third column in (6.29) by γ , and adding it to the fourth column, we obtain,

$$|\underline{p}_1, \tilde{R}^* \underline{p}_1, \underline{q}_2, \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4.$$

Next, multiplying the first column by $-\bar{\lambda}_1$, and adding it to the second column results in,

$$|\underline{p}_1, (\tilde{R}^* - \bar{\lambda}_1 I) \underline{p}_1, \underline{q}_2, \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4.$$

If the vectors, \underline{p}_1 , $(\tilde{R}^* - \bar{\lambda}_1 I)\underline{p}_1$, and \underline{q}_2 , were linearly independent, we could always choose $\underline{w} \in \mathcal{C}^4$ so that

$$|\underline{p}_1, (\tilde{R}^* - \bar{\lambda}_1 I)\underline{p}_1, \underline{q}_2, \underline{w}| \neq 0.$$

Therefore, \underline{p}_1 , $(\tilde{R}^* - \bar{\lambda}_1 I)\underline{p}_1$, and \underline{q}_2 must be linearly dependent. This implies that

$$(\tilde{R}^* - \bar{\lambda}_1 I)\underline{p}_1 \in \text{sp}\{\underline{p}_1, \underline{q}_2\}. \quad (6.30)$$

Next, by choosing $\underline{y} = \underline{q}_3$ in (6.28), where $\tilde{R}^*\underline{q}_3 = \bar{\lambda}_3\underline{q}_3$, and repeating the above argument, we obtain

$$(\tilde{R}^* - \bar{\lambda}_1 I)\underline{p}_1 \in \text{sp}\{\underline{p}_1, \underline{q}_3\}. \quad (6.31)$$

Since \underline{p}_1 is orthogonal to both \underline{q}_2 and \underline{q}_3 , and $\underline{q}_2, \underline{q}_3$ are linearly independent,

$$(\tilde{R}^* - \bar{\lambda}_1 I)\underline{p}_1 \in \text{sp}\{\underline{p}_1, \underline{q}_2\} \cap \text{sp}\{\underline{p}_1, \underline{q}_3\} = \text{sp}\{\underline{p}_1\}.$$

Thus, $(\tilde{R}^* - \bar{\lambda}_1 I)\underline{p}_1 = \gamma\underline{p}_1$, for some γ , which implies \underline{p}_1 is an eigenvector of \tilde{R}^* . We have established that \underline{p}_1 is an eigenvector of both \tilde{R} and \tilde{R}^* .

If we begin by choosing \underline{p}_j to be one of the other eigenvectors in (6.25), the same argument can be repeated to show that the other eigenvectors, \underline{p}_j of \tilde{R} with eigenvalue λ_j , are also eigenvectors of \tilde{R}^* with eigenvalue $\bar{\lambda}_j$. Thus, \tilde{R} and \tilde{R}^* have the same complete set of eigenvectors, and \tilde{R} is normal on \mathcal{C}^4 .

From (6.18), we see that if \underline{p}_j is an eigenvector of \tilde{R} and \tilde{R}^* ,

$$\tilde{R}\underline{p}_j = \tilde{Q}^{-1}\tilde{A}\tilde{Q}\underline{p}_j = \lambda_j\underline{p}_j,$$

and thus

$$\tilde{A}(\tilde{Q}\underline{p}_j) = \lambda_j(\tilde{Q}\underline{p}_j).$$

Therefore, $\tilde{Q}\underline{p}_j$ is an eigenvector of \tilde{A} . Using (6.18), we obtain

$$\begin{aligned} \tilde{A}^\dagger\tilde{Q}\underline{p}_j &= B^{-1}(\tilde{Q}^{-1})^*\tilde{R}^*\tilde{Q}^*B(\tilde{Q}\underline{p}_j) = \tilde{Q}\tilde{R}^*\tilde{Q}^{-1}(\tilde{Q}\underline{p}_j) \\ &= \tilde{Q}\tilde{R}^*\underline{p}_j = \bar{\lambda}_j\tilde{Q}\underline{p}_j. \end{aligned}$$

It follows that $\tilde{Q}\underline{p}_j$ is also an eigenvector of \tilde{A}^\dagger with eigenvalue $\bar{\lambda}_j$. Therefore, \tilde{A} and \tilde{A}^\dagger have the same complete set of eigenvectors. Finally, notice that if $\tilde{Q}\underline{p}_j$ and $\tilde{Q}\underline{p}_k$ are any two eigenvectors of \tilde{A} and \tilde{A}^\dagger ,

$$\langle B\tilde{Q}\underline{p}_j, \tilde{Q}\underline{p}_k \rangle = \langle \tilde{Q}^* B\tilde{Q}\underline{p}_j, \underline{p}_k \rangle = \langle \underline{p}_j, \underline{p}_k \rangle = 0,$$

which means \tilde{A} and \tilde{A}^\dagger have the same complete set of B -orthogonal eigenvectors, thus, \tilde{A} is B -normal of \mathcal{V}_4 .

Every eigenvector of A can be included in a 4-dimensional invariant subspace of A , upon which the restriction of A is B -normal. It follows that each eigenvector of A with λ_j , is also an eigenvector of A^\dagger with $\bar{\lambda}_j$. Furthermore, the eigenvectors are B -orthonormal, thus, A is B -normal.

Since A is B -normal, it has a diagonal Schur form ([22], pp. 329), that is,

$$A = QRQ^{-1}, \quad Q^* BQ = I \in \mathcal{C}^{N \times N}, \quad (6.32)$$

where R is a diagonal matrix, whose entries are the eigenvalues of A , and Q is B -orthonormal.

Consider the wedge product,

$$\mathcal{F}(\underline{w}, R) = \underline{w} \wedge R\underline{w} \wedge R^*\underline{w} \wedge R^*R\underline{w},$$

where

$$R = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_N \end{bmatrix},$$

and \underline{w} is any vector in \mathcal{C}^N . Denote,

$$\underline{w} = [\eta_1, \eta_2, \dots, \eta_N]^T,$$

and let

$$\Lambda = [\underline{w}, R\underline{w}, R^*\underline{w}, R^*R\underline{w}] = \begin{bmatrix} \eta_1 & \lambda_1\eta_1 & \bar{\lambda}_1\eta_1 & \bar{\lambda}_1\lambda_1\eta_1 \\ \eta_2 & \lambda_2\eta_2 & \bar{\lambda}_2\eta_2 & \bar{\lambda}_2\lambda_2\eta_2 \\ \vdots & \vdots & \vdots & \vdots \\ \eta_N & \lambda_N\eta_N & \bar{\lambda}_N\eta_N & \bar{\lambda}_N\lambda_N\eta_N \end{bmatrix}. \quad (6.33)$$

Recall that $\mathcal{F}(\underline{w}, R)$ is a mapping from

$$\mathcal{C}^{N \times 4} \longrightarrow \mathcal{C}^{\binom{N}{4}}.$$

\mathcal{F} takes the matrix Λ , and maps it onto a vector \underline{f} of length $\binom{N}{4}$. Each element of \underline{f} corresponds to choosing 4 of the N rows of Λ , and taking the determinant of it. If all $\binom{N}{4}$ determinants are zero, then $\mathcal{F}(\underline{w}, R) \equiv \underline{0}$. It follows from the previous steps, that each element of \underline{f} is equivalent to $\mathcal{F}(\underline{\tilde{w}}, \tilde{R})$, for some $\underline{\tilde{w}} \in \mathcal{C}^4$. From Lemma 6.7 and the discussion following it, $\mathcal{F}(\underline{\tilde{w}}, \tilde{R})$ was shown to be zero for all $\underline{\tilde{w}} \in \mathcal{C}^4$. Therefore, $\mathcal{F}(\underline{w}, R) \equiv \underline{0}$, for all \underline{w} in \mathcal{C}^N . This means there exists constants, α_0 , α_1 , β_0 , and β_1 , not all zero, satisfying

$$\alpha_0\underline{w} + \alpha_1 R\underline{w} + \beta_0 R^*\underline{w} + \beta_1 R^*R\underline{w} = \underline{0}. \quad (6.34)$$

Equivalently, using (6.33), we obtain

$$\begin{bmatrix} \eta_1 & \lambda_1\eta_1 & \bar{\lambda}_1\eta_1 & \bar{\lambda}_1\lambda_1\eta_1 \\ \eta_2 & \lambda_2\eta_2 & \bar{\lambda}_2\eta_2 & \bar{\lambda}_2\lambda_2\eta_2 \\ \vdots & \vdots & \vdots & \vdots \\ \eta_N & \lambda_N\eta_N & \bar{\lambda}_N\eta_N & \bar{\lambda}_N\lambda_N\eta_N \end{bmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_0 \\ \beta_1 \end{pmatrix} = \underline{0}. \quad (6.35)$$

Next, we show that these constants can be chosen independent of the vector \underline{w} . Notice that we can factor \underline{w} out of (6.35), producing

$$\begin{bmatrix} \eta_1 & & & \\ & \eta_2 & & \\ & & \ddots & \\ & & & \eta_N \end{bmatrix} \begin{bmatrix} 1 & \lambda_1 & \bar{\lambda}_1 & \bar{\lambda}_1 \lambda_1 \\ 1 & \lambda_2 & \bar{\lambda}_2 & \bar{\lambda}_2 \lambda_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_N & \bar{\lambda}_N & \bar{\lambda}_N \lambda_N \end{bmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_0 \\ \beta_1 \end{pmatrix} = \underline{0}. \quad (6.36)$$

Since (6.36) must hold for every \underline{w} , it must hold for

$$\underline{w} = [\eta_1, \eta_2, \dots, \eta_N]^T = [1, 1, \dots, 1]^T.$$

This implies that α_0 , α_1 , β_0 , and β_1 can be chosen to satisfy

$$\begin{bmatrix} 1 & \lambda_1 & \bar{\lambda}_1 & \bar{\lambda}_1 \lambda_1 \\ 1 & \lambda_2 & \bar{\lambda}_2 & \bar{\lambda}_2 \lambda_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_N & \bar{\lambda}_N & \bar{\lambda}_N \lambda_N \end{bmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_0 \\ \beta_1 \end{pmatrix} = \underline{0}. \quad (6.37)$$

If α_0 , α_1 , β_0 , and β_1 , are chosen to satisfy (6.37), then they also satisfy (6.36), for any vector $\underline{w} \in \mathcal{C}^N$. Thus, these constants can be chosen independent of \underline{w} . It follows from (6.34) that for every \underline{w} ,

$$R^*[\beta_1 R + \beta_0 I]\underline{w} = -[\alpha_1 R + \alpha_0 I]\underline{w},$$

or equivalently,

$$R^* q_1(R)\underline{w} = p_1(R)\underline{w},$$

where $p_1(\lambda)$ and $q_1(\lambda)$ each have degree at most 1, and can be chosen the same for every \underline{w} . This implies that

$$R^* q_1(R) = p_1(R). \quad (6.38)$$

It follows from Definition 4.1, that R is I -normal(ℓ, m), with $(\ell, m) \leq (1, 1)$.

We will now show that this implies that A is B -normal(ℓ, m) with $(\ell, m) \leq (1, 1)$. From (6.32) we obtain,

$$R^* = Q^*A^*(Q^{-1})^*, \quad \text{and} \quad Q^*B = Q^{-1}.$$

Since $A^\dagger = B^{-1}A^*B$, we can write,

$$A^* = BA^\dagger B^{-1}.$$

Using this information to expand (6.38) yields,

$$Q^{-1}A^\dagger q_1(A)Q = Q^{-1}p_1(A)Q,$$

and thus

$$A^\dagger q_1(A) = p_1(A),$$

for some polynomials, p_1 and q_1 , each of degree at most 1. It follows from Definition 4.1, that A is B -normal(ℓ, m), with $(\ell, m) \leq (1, 1)$. The proof for the diagonalizable case is complete.

Part B: Suppose that A is not diagonalizable.

We have separated this case from the main proof due to the fact that this proof for ruling out nonlinear elementary divisors is a rather lengthy case argument. It is included for completeness. We recommend that only the most ambitious readers need proceed.

We will show that if $A \in \mathcal{CG}_{\text{SR}}(1, 1)$, and $d(A) > 3$, then A cannot have any nonlinear elementary divisors. This will complete the proof of the necessary conditions.

Suppose $A \in \mathcal{CG}_{\text{SR}}(1, 1)$, and $d(A) > 3$. Choose \underline{p}_0 such that $d(\underline{p}_0, A) = 4$, and let \tilde{A} to be the restriction of A to \mathcal{V}_4 , where,

$$\mathcal{V}_4 = \text{sp}\{\underline{p}_0, A\underline{p}_0, A^2\underline{p}_0, A^3\underline{p}_0\}.$$

Recall that the relationships given in (6.24) and (6.25) hold, where \tilde{R} denotes the upper triangular, matrix representation that results from Schur's Theorem, and $G_j = (\tilde{R} - \lambda_j I)$. We begin by showing the \tilde{R} does not have any nonlinear elementary divisors.

Case 1: Suppose the minimal polynomial of \tilde{R} is given by

$$p_4(\lambda) = (\lambda - \lambda_1)^4,$$

that is, \tilde{R} has a nonlinear elementary divisor of size 4. Let $\underline{p}_{1_1}, (\underline{p}_{1_2}, \underline{p}_{1_3}, \underline{p}_{1_4})$ be the eigenvector and generalized eigenvectors of \tilde{R} associated with eigenvalue λ_1 . Similarly, denote $\underline{q}_{1_1}, (\underline{q}_{1_2}, \underline{q}_{1_3}, \underline{q}_{1_4})$, as the eigenvector and generalized eigenvectors of \tilde{R}^* associated with eigenvalue $\bar{\lambda}_1$. Recall that the eigenvectors and generalized eigenvectors of \tilde{R} and \tilde{R}^* can be chosen to satisfy the following relationships:

$$\begin{aligned} (\tilde{R} - \lambda_1 I)\underline{p}_{1_1} &= \underline{0} & (\tilde{R}^* - \bar{\lambda}_1 I)\underline{q}_{1_1} &= \underline{0} \\ (\tilde{R} - \lambda_1 I)\underline{p}_{1_2} &= \underline{p}_{1_1} & (\tilde{R}^* - \bar{\lambda}_1 I)\underline{q}_{1_2} &= \underline{q}_{1_1} \\ (\tilde{R} - \lambda_1 I)\underline{p}_{1_3} &= \underline{p}_{1_2} & (\tilde{R}^* - \bar{\lambda}_1 I)\underline{q}_{1_3} &= \underline{q}_{1_2} \\ (\tilde{R} - \lambda_1 I)\underline{p}_{1_4} &= \underline{p}_{1_3} & (\tilde{R}^* - \bar{\lambda}_1 I)\underline{q}_{1_4} &= \underline{q}_{1_3} \end{aligned} \tag{6.39}$$

Notice that \underline{p}_{1_4} is not in the range of G_1 , and \underline{q}_{1_4} is not in the range of G_1^* . Recall that the Jordan decomposition of \tilde{R} is given by

$$\tilde{R} = SJS^{-1}, \quad J = \begin{bmatrix} \lambda_1 & 1 & & \\ & \lambda_1 & 1 & \\ & & \lambda_1 & 1 \\ & & & \lambda_1 \end{bmatrix},$$

where the columns of S are the eigenvectors and generalized eigenvectors of \tilde{R} . Since,

$$\tilde{R}^* = (S^{-1})^* J^* S^*,$$

it follows that the columns of $(S^{-1})^*$ (conjugates of the rows of S^{-1}) are the eigenvectors and generalized eigenvectors of \tilde{R}^* . Notice that the order of the columns of $(S^{-1})^*$ is given by, $(S^{-1})^* = [\underline{q}_{14}, \underline{q}_{13}, \underline{q}_{12}, \underline{q}_{11}]$, since (6.39) implies that

$$\tilde{R}^* [\underline{q}_{14}, \underline{q}_{13}, \underline{q}_{12}, \underline{q}_{11}] = [\underline{q}_{14}, \underline{q}_{13}, \underline{q}_{12}, \underline{q}_{11}] \begin{bmatrix} \bar{\lambda}_1 & & & \\ & 1 & \bar{\lambda}_1 & \\ & & 1 & \bar{\lambda}_1 \\ & & & 1 & \bar{\lambda}_1 \end{bmatrix}.$$

Furthermore, we see that the following orthogonality relationships hold:

$$S^{-1}S = \begin{bmatrix} - & \underline{q}_{14}^* & - \\ - & \underline{q}_{13}^* & - \\ - & \underline{q}_{12}^* & - \\ - & \underline{q}_{11}^* & - \end{bmatrix} \begin{bmatrix} | & | & | & | \\ \underline{p}_{11} & \underline{p}_{12} & \underline{p}_{13} & \underline{p}_{14} \\ | & | & | & | \end{bmatrix} = I, \quad (6.40)$$

where, I is the 4×4 identity matrix. We see that \underline{p}_{11} is orthogonal to \underline{q}_{13} , \underline{q}_{12} , and \underline{q}_{11} . In general, each eigenvector (generalized eigenvector), \underline{p}_{1j} of \tilde{R} , given by the j 'th column of S , is orthogonal to all but the eigenvector (generalized eigenvector) of \tilde{R}^* , whose conjugate occupies the j 'th row of S^{-1} .

From (6.24), with $\underline{p}_j = \underline{p}_{11}$, and $\lambda_j = \lambda_1$, it follows that

$$|\underline{p}_{11}, G_1^* \underline{p}_{11}, G_1 \underline{v}, G_1^* G_1 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4. \quad (6.41)$$

The range of G_1 is orthogonal to the nullspace of G_1^* . Using (6.39), we see that G_1^* has a one dimensional nullspace spanned by \underline{q}_{11} . Choose $\underline{v} \in \mathcal{C}^4$ such that

$$G_1 \underline{v} = \underline{q}_{12} + \gamma \underline{q}_{11}, \quad (6.42)$$

where γ is chosen such that $G_1 \underline{v}$ is orthogonal to \underline{q}_{11} , that is, $\gamma = -\frac{\langle \underline{q}_{12}, \underline{q}_{11} \rangle}{\langle \underline{q}_{11}, \underline{q}_{11} \rangle}$. It follows from (6.39) that,

$$G_1^* G_1 \underline{v} = \underline{q}_{11}. \quad (6.43)$$

We remark here, that (6.43) makes sense because \underline{q}_{1_1} is in the range of $G_1^*G_1$. This is justified by noting that G_1 has a one dimensional nullspace spanned by \underline{p}_{1_1} . Since the nullspace of G_1 and $G_1^*G_1$ are equivalent, the nullspace of $G_1^*G_1$ is spanned by \underline{p}_{1_1} . Furthermore, the nullspace of $G_1^*G_1$ is orthogonal to the range of $(G_1^*G_1)^*$, which equals the range of $G_1^*G_1$. Therefore, the range of $G_1^*G_1$ is the set of vectors in \mathcal{C}^4 that are orthogonal to $\text{sp}\{\underline{p}_{1_1}\}$. From the above orthogonality relationships, $\underline{q}_{1_1} - \underline{p}_{1_1}$, and thus \underline{q}_{1_1} is in the range of $G_1^*G_1$.

Substituting (6.42) and (6.43) into (6.41) yields,

$$|\underline{p}_{1_1}, G_1^*\underline{p}_{1_1}, \underline{q}_{1_2} + \gamma\underline{q}_{1_1}, \underline{q}_{1_1}| = 0.$$

Next, by using the properties of determinants to expand and then delete terms with dependent columns, we obtain,

$$|\underline{p}_{1_1}, G_1^*\underline{p}_{1_1}, \underline{q}_{1_2}, \underline{q}_{1_1}| = 0. \quad (6.44)$$

Since \underline{p}_{1_1} is in the nullspace of G_1 , it follows that $\langle \underline{p}_{1_1}, G_1^*\underline{p}_{1_1} \rangle = \langle G_1\underline{p}_{1_1}, \underline{p}_{1_1} \rangle = 0$. From (6.40), we see that \underline{p}_{1_1} is also orthogonal to \underline{q}_{1_1} , and \underline{q}_{1_2} . In order to satisfy (6.44), this means that the vectors, $G_1^*\underline{p}_{1_1}$, \underline{q}_{1_1} , and \underline{q}_{1_2} must be linearly dependent, or

$$\alpha G_1^*\underline{p}_{1_1} + \beta \underline{q}_{1_1} + \gamma \underline{q}_{1_2} = \underline{0},$$

for some constants, α , β , and γ , not all zero. If $\alpha = 0$, this would imply that \underline{q}_{1_1} and \underline{q}_{1_2} were linearly dependent, which is a contradiction. Therefore,

$$G_1^*\underline{p}_{1_1} \in \text{sp}\{\underline{q}_{1_1}, \underline{q}_{1_2}\}.$$

If $G_1^*\underline{p}_{1_1} = \underline{0}$, then \underline{p}_{1_1} would also be an eigenvector of \tilde{R}^* with eigenvalue $\bar{\lambda}_1$. This leads to a contradiction since we have assumed that \underline{q}_{1_1} is the only eigenvector of \tilde{R}^* with eigenvalue $\bar{\lambda}_1$, and using (6.40), we see that $\underline{p}_{1_1} - \underline{q}_{1_1}$. Suppose that $G_1^*\underline{p}_{1_1} \neq \underline{0}$,

and notice from (6.40), that \underline{p}_{1_2} is orthogonal to $\text{sp}\{\underline{q}_{1_1}, \underline{q}_{1_2}\}$. It follows that \underline{p}_{1_2} must be orthogonal to $G_1^* \underline{p}_{1_1}$. This means,

$$0 = \langle G_1^* \underline{p}_{1_1}, \underline{p}_{1_2} \rangle = \langle \underline{p}_{1_1}, G_1 \underline{p}_{1_2} \rangle = \langle \underline{p}_{1_1}, \underline{p}_{1_1} \rangle.$$

However, since the inner product is definite,

$$\langle \underline{p}_{1_1}, \underline{p}_{1_1} \rangle = 0 \iff \underline{p}_{1_1} \equiv \underline{0}.$$

This is a contradiction to our assumption that \underline{p}_{1_1} is an eigenvector of \tilde{R} . Therefore, the minimal polynomial of \tilde{R} cannot be given by

$$p_4(\lambda) = (\lambda - \lambda_1)^4.$$

Case 2: Suppose the minimal polynomial of \tilde{R} is given by

$$p_4(\lambda) = (\lambda - \lambda_1)^3(\lambda - \lambda_2).$$

Denote $\underline{p}_{1_1}, (\underline{p}_{1_2}, \underline{p}_{1_3})$ as the eigenvector and generalized eigenvectors of \tilde{R} associated with eigenvalue λ_1 , and \underline{p}_{2_1} as the eigenvector of \tilde{R} with eigenvalue λ_2 . Similarly, denote $\underline{q}_{1_1}, (\underline{q}_{1_2}, \underline{q}_{1_3})$ as the eigenvector and generalized eigenvectors of \tilde{R}^* associated with eigenvalue $\bar{\lambda}_1$, and \underline{q}_{2_1} as the eigenvector of \tilde{R}^* with eigenvalue $\bar{\lambda}_2$. Recall the following relationships between the eigenvectors and generalized eigenvectors of \tilde{R} and \tilde{R}^* :

$$\begin{aligned} (\tilde{R} - \lambda_1 I) \underline{p}_{1_1} &= \underline{0} & (\tilde{R}^* - \bar{\lambda}_1 I) \underline{q}_{1_1} &= \underline{0} \\ (\tilde{R} - \lambda_1 I) \underline{p}_{1_2} &= \underline{p}_{1_1} & (\tilde{R}^* - \bar{\lambda}_1 I) \underline{q}_{1_2} &= \underline{q}_{1_1} \\ (\tilde{R} - \lambda_1 I) \underline{p}_{1_3} &= \underline{p}_{1_2} & (\tilde{R}^* - \bar{\lambda}_1 I) \underline{q}_{1_3} &= \underline{q}_{1_2} \\ (\tilde{R} - \lambda_2 I) \underline{p}_{2_1} &= \underline{0} & (\tilde{R}^* - \bar{\lambda}_2 I) \underline{q}_{2_1} &= \underline{0} \end{aligned} \tag{6.45}$$

From the Jordan decompositions of \tilde{R} and \tilde{R}^* ,

$$\tilde{R} = SJS^{-1}, \quad \tilde{R}^* = (S^{-1})^* J^* S^*,$$

the following orthogonality condition can be derived:

$$S^{-1}S = \begin{bmatrix} - & \underline{q}_{13}^* & - \\ - & \underline{q}_{12}^* & - \\ - & \underline{q}_{11}^* & - \\ - & \underline{q}_{21}^* & - \end{bmatrix} \begin{bmatrix} | & | & | & | \\ \underline{p}_{11} & \underline{p}_{12} & \underline{p}_{13} & \underline{p}_{21} \\ | & | & | & | \end{bmatrix} = I. \quad (6.46)$$

Substituting $\underline{p}_j = \underline{p}_{21}$, and $\lambda_j = \lambda_2$ into (6.24), shows that the following condition must hold:

$$|\underline{p}_{21}, G_2^* \underline{p}_{21}, G_2 \underline{v}, G_2^* G_2 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4. \quad (6.47)$$

Since $\mathcal{R}(G_2)$ is orthogonal to $\mathcal{N}(G_2)^* = \text{sp}\{\underline{q}_{21}\}$, it follows from (6.46) that $\mathcal{R}(G_2)$ is spanned by $\{\underline{p}_{11}, \underline{p}_{12}, \underline{p}_{13}\}$. Any vector $\underline{w} \in \mathcal{C}^4$ can be written as $G_2 \underline{v} + \gamma \underline{p}_{21}$, for some $\underline{v} \in \mathcal{C}^4$ and some $\gamma \in \mathcal{C}$. By multiplying the first column in (6.47) by γ and adding it to the third column, then multiplying the second column by γ and adding to the forth column produces

$$|\underline{p}_{21}, G_2^* \underline{p}_{21}, \underline{w}, G_2^* \underline{w}|, \quad \forall \underline{w} \in \mathcal{C}^4.$$

Next, we multiply the third column by $(\bar{\lambda}_2 - \bar{\lambda}_1)I$ and add it to the forth column which yields

$$|\underline{p}_{21}, G_2^* \underline{p}_{21}, \underline{w}, G_1^* \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4. \quad (6.48)$$

Let $\underline{w} = \underline{q}_{12}$, then using (6.45) we obtain that $G_1^* \underline{w} = \underline{q}_{11}$. Substituting these into the above results in

$$|\underline{p}_{21}, G_2^* \underline{p}_{21}, \underline{q}_{12}, \underline{q}_{11}| = 0.$$

Since the set $\{\underline{p}_{21}, \underline{q}_{12}, \underline{q}_{11}\}$ is linearly independent, in order for the above determinant to be zero requires that $G_2^* \underline{p}_{21} \in \text{sp}\{\underline{p}_{21}, \underline{q}_{12}, \underline{q}_{11}\}$. Furthermore, notice that $\underline{p}_{21} \in \text{sp}\{G_2^* \underline{p}_{21}, \underline{q}_{12}, \underline{q}_{11}\}$, and it follows that

$$G_2^* \underline{p}_{21} \in \text{sp}\{\underline{q}_{12}, \underline{q}_{11}\}. \quad (6.49)$$

Next, let $\underline{w} = \underline{q}_{13}$, and use (6.45) to show that $G_1^* \underline{w} = \underline{q}_{12}$. Substituting these expressions into (6.48) yields

$$|\underline{p}_{21}, G_2^* \underline{p}_{21}, \underline{q}_{13}, \underline{q}_{12}| = 0.$$

Since the set $\{\underline{p}_{21}, \underline{q}_{13}, \underline{q}_{12}\}$ is linearly independent, and $\underline{p}_{21} \in \text{sp}\{G_2^* \underline{p}_{21}, \underline{q}_{13}, \underline{q}_{12}\}$, it follows that

$$G_2^* \underline{p}_{21} \in \text{sp}\{\underline{q}_{13}, \underline{q}_{12}\}. \quad (6.50)$$

Together, (6.49) and (6.50) yield

$$\begin{aligned} G_2^* \underline{p}_{21} &\in \text{sp}\{\underline{q}_{12}, \underline{q}_{11}\} \cap \text{sp}\{\underline{q}_{13}, \underline{q}_{12}\} \\ &= \text{sp}\{\underline{q}_{12}\}, \end{aligned}$$

thus,

$$G_2^* \underline{p}_{21} = \alpha \underline{q}_{12}.$$

Suppose that $G_2^* \underline{p}_{21} = \alpha \underline{q}_{12}$ for some $\alpha \neq 0$. Plugging this into (6.48) and using the properties of determinants to simplify produces

$$\alpha |\underline{p}_{21}, \underline{q}_{12}, \underline{w}, G_1^* \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4.$$

Next, substituting $\underline{w} = \underline{q}_{13} + \underline{q}_{12}$ into the above, then using (6.45) and the properties of determinants to expand and delete those terms with dependent columns yields

$$\alpha |\underline{p}_{21}, \underline{q}_{12}, \underline{q}_{13}, \underline{q}_{11}| = 0.$$

Since the vectors $\underline{p}_{21}, \underline{q}_{12}, \underline{q}_{13}$, and \underline{q}_{11} are linearly independent, it follows that $\alpha = 0$, which means $G_2^* \underline{p}_{21} = \underline{0}$.

If $G_2^* \underline{p}_{21} = \underline{0}$, \underline{p}_{21} is an eigenvector of R^* with eigenvalue $\bar{\lambda}_2$, which means $\underline{p}_{21} \in \text{sp}\{\underline{q}_{21}\}$. From the orthogonality condition given in (6.46), we see that $\underline{p}_{21} \in \text{sp}\{\underline{p}_{11}, \underline{p}_{12}, \underline{p}_{13}\}$. Notice that this means the matrices \tilde{R} and G_2 have the forms

$$\tilde{R} = \begin{bmatrix} \lambda_1 & r_{1,2} & r_{1,3} & 0 \\ & \lambda_1 & r_{2,3} & 0 \\ & & \lambda_1 & 0 \\ & & & \lambda_2 \end{bmatrix}, \quad G_2 = \begin{bmatrix} \lambda_1 - \lambda_2 & r_{1,2} & r_{1,3} & 0 \\ & \lambda_1 - \lambda_2 & r_{2,3} & 0 \\ & & \lambda_1 - \lambda_2 & 0 \\ & & & 0 \end{bmatrix}.$$

Recall from (6.22) with $\lambda_j = \lambda_2$, that

$$|\tilde{\underline{w}}, G_2\tilde{\underline{w}}, G_2^*\tilde{\underline{w}}, G_2^*G_2\tilde{\underline{w}}| = 0, \quad \forall \tilde{\underline{w}} \in \mathcal{C}^4.$$

Next, writing $\tilde{\underline{w}} = \underline{u} + \alpha\underline{v}$, for some $\underline{u}, \underline{v} \in \mathcal{C}^4$ and some $\alpha \in \mathcal{C}$, the above becomes

$$|\underline{u} + \alpha\underline{v}, G_2(\underline{u} + \alpha\underline{v}), G_2^*(\underline{u} + \alpha\underline{v}), G_2^*G_2(\underline{u} + \alpha\underline{v})| = 0,$$

for every $\underline{u}, \underline{v} \in \mathcal{C}^4$ and every $\alpha \in \mathcal{C}$. By differentiating once with respect to α , then setting $\alpha = 0$, we obtain:

$$\begin{aligned} & |\underline{v}, G_2\underline{u}, G_2^*\underline{u}, G_2^*G_2\underline{u}| + |\underline{u}, G_2\underline{v}, G_2^*\underline{u}, G_2^*G_2\underline{u}| \\ & |\underline{u}, G_2\underline{u}, G_2^*\underline{v}, G_2^*G_2\underline{u}| + |\underline{u}, G_2\underline{u}, G_2^*\underline{u}, G_2^*G_2\underline{v}| = 0, \end{aligned}$$

for every $\underline{u}, \underline{v} \in \mathcal{C}^4$. Since \underline{p}_{2_1} is an eigenvector of both \tilde{R} and \tilde{R}^* , by letting $\underline{v} = \underline{p}_{2_1}$ in the above, all the determinants vanish except for the first one, yielding

$$|\underline{p}_{2_1}, G_2\underline{u}, G_2^*\underline{u}, G_2^*G_2\underline{u}| = 0, \quad \forall \underline{u} \in \mathcal{C}^4.$$

Notice that $\underline{p}_{2_1} - \text{sp}\{G_2\underline{u}, G_2^*\underline{u}, G_2^*G_2\underline{u}\}$ which means the vectors $G_2\underline{u}$, $G_2^*\underline{u}$ and $G_2^*G_2\underline{u}$ must be linearly dependent, or equivalently,

$$G_2\underline{u} \wedge G_2^*\underline{u} \wedge G_2^*G_2\underline{u} = 0, \quad \forall \underline{u} \in \mathcal{C}^4. \quad (6.51)$$

Denote

$$\hat{G}_2 = \begin{bmatrix} \lambda_1 - \lambda_2 & r_{1,2} & r_{1,3} \\ & \lambda_1 - \lambda_2 & r_{2,3} \\ & & \lambda_1 - \lambda_2 \end{bmatrix} \quad \hat{R} = \begin{bmatrix} \lambda_1 & r_{1,2} & r_{1,3} \\ & \lambda_1 & r_{2,3} \\ & & \lambda_1 \end{bmatrix}.$$

Notice that if \tilde{R} has a nonlinear elementary divisor of size three associated with λ_1 , so does \hat{R} . Furthermore, it follows that \hat{G}_2 has a nonlinear elementary divisor of size

three associated with $\lambda_1 - \lambda_2$. Since the matrices G_2 and G_2^* have all zeros in their last row and column, (6.51) holds if and only if

$$\hat{G}_2 \underline{z} \wedge \hat{G}_2^* \underline{z} \wedge \hat{G}_2^* \hat{G}_2 \underline{z} = |\hat{G}_2 \underline{z}, \hat{G}_2^* \underline{z}, \hat{G}_2 \hat{G}_2 \underline{z}| = 0,$$

for every $\underline{z} \in \mathcal{C}^3$. Notice that \hat{G}_2 is a nonsingular upper triangular matrix. The above determinant is still zero with any change of basis. Let $\underline{z} = \hat{G}_2^{-1} \underline{w}$ and then multiply through on left by $(\hat{G}_2^{-1})^*$ to obtain

$$|(\hat{G}_2^{-1})^* \underline{w}, \hat{G}_2^{-1} \underline{w}, \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^3.$$

By denoting $\hat{G} = \hat{G}_2^{-1}$ and rearranging, we obtain

$$|\underline{w}, \hat{G} \underline{w}, \hat{G}^* \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4. \quad (6.52)$$

Let

$$\{\hat{\underline{p}}_{1_1}, \hat{\underline{p}}_{1_2}, \hat{\underline{p}}_{1_3}\} \quad \text{and} \quad \{\hat{\underline{q}}_{1_1}, \hat{\underline{q}}_{1_2}, \hat{\underline{q}}_{1_3}\},$$

denote the eigenvectors and generalized eigenvectors of \hat{G} and \hat{G}^* respectively. Next, set

$$\underline{w} = \hat{\underline{p}}_{1_2} + \alpha \hat{\underline{p}}_{1_1},$$

where α is chosen so that

$$\langle \underline{w}, \hat{\underline{p}}_{1_1} \rangle = 0.$$

Since

$$\begin{aligned} \langle \hat{G}^* \underline{w}, \hat{\underline{p}}_{1_1} \rangle &= \langle \underline{w}, \hat{G} \hat{\underline{p}}_{1_1} \rangle = \langle \underline{w}, \underline{0} \rangle = 0, \quad \text{and} \\ \langle \hat{G}^* \underline{w}, \hat{\underline{p}}_{1_2} \rangle &= \langle \underline{w}, \hat{G} \hat{\underline{p}}_{1_2} \rangle = \langle \underline{w}, \hat{\underline{p}}_{1_1} \rangle = 0, \end{aligned}$$

it follows that $\hat{G}^* \underline{w} = \beta \hat{\underline{q}}_{1_1}$, for some β . Substituting this information into (6.52) yields,

$$|\hat{\underline{p}}_{1_2} + \alpha \hat{\underline{p}}_{1_1}, \hat{\underline{p}}_{1_1}, \beta \hat{\underline{q}}_{1_1}| = \beta |\hat{\underline{p}}_{1_2}, \hat{\underline{p}}_{1_1}, \hat{\underline{q}}_{1_1}| = 0.$$

Since $\hat{\underline{p}}_{1_2}$, $\hat{\underline{p}}_{1_1}$, and $\hat{\underline{q}}_{1_1}$ are linearly independent, it follows that $\beta = 0$. This implies $\hat{\underline{p}}_{1_2} + \alpha \hat{\underline{p}}_{1_1}$ is an eigenvector of \hat{R}^* that is orthogonal to $\hat{\underline{q}}_{1_1}$, which is a contradiction.

It follows that \tilde{R} cannot have a nonlinear elementary divisor of size three, thus, the minimal polynomial of \tilde{R} is not given by,

$$p_4(\lambda) = (\lambda - \lambda_1)^3(\lambda - \lambda_2).$$

Case 3: Suppose the minimal polynomial of \tilde{R} is given by

$$p_4(\lambda) = (\lambda - \lambda_1)^2(\lambda - \lambda_2)^2.$$

Denote \underline{p}_{1_1} (\underline{p}_{1_2}), as the eigenvector (generalized eigenvector) of \tilde{R} associated with eigenvalue λ_1 , and \underline{p}_{2_1} (\underline{p}_{2_2}), as the eigenvector (generalized eigenvector) of \tilde{R} with eigenvalue λ_2 . Similarly, \underline{q}_{1_1} (\underline{q}_{1_2}), denotes the eigenvector (generalized eigenvector) of \tilde{R}^* with eigenvalue $\bar{\lambda}_1$, and \underline{q}_{2_1} (\underline{q}_{2_2}), are the eigenvector and generalized eigenvector with eigenvalue $\bar{\lambda}_2$. The following relationships hold between the eigenvectors and generalized eigenvectors of \tilde{R} and \tilde{R}^* :

$$\begin{aligned} (\tilde{R} - \lambda_1 I)\underline{p}_{1_1} &= \underline{0} & (\tilde{R}^* - \bar{\lambda}_1 I)\underline{q}_{1_1} &= \underline{0} \\ (\tilde{R} - \lambda_1 I)\underline{p}_{1_2} &= \underline{p}_{1_1} & (\tilde{R}^* - \bar{\lambda}_1 I)\underline{q}_{1_2} &= \underline{q}_{1_1} \\ (\tilde{R} - \lambda_2 I)\underline{p}_{2_1} &= \underline{0} & (\tilde{R}^* - \bar{\lambda}_2 I)\underline{q}_{2_1} &= \underline{0} \\ (\tilde{R} - \lambda_2 I)\underline{p}_{2_2} &= \underline{p}_{2_1} & (\tilde{R}^* - \bar{\lambda}_2 I)\underline{q}_{2_2} &= \underline{q}_{2_1} \end{aligned} \tag{6.53}$$

It follows from the Jordan decompositions of \tilde{R} , and \tilde{R}^* , that

$$S^{-1}S = \begin{bmatrix} - & \underline{q}_{1_2}^* & - \\ - & \underline{q}_{1_1}^* & - \\ - & \underline{q}_{2_2}^* & - \\ - & \underline{q}_{2_1}^* & - \end{bmatrix} \begin{bmatrix} | & | & | & | \\ \underline{p}_{1_1} & \underline{p}_{1_2} & \underline{p}_{2_1} & \underline{p}_{2_2} \\ | & | & | & | \end{bmatrix} = I. \tag{6.54}$$

By Substituting $\underline{p}_j = \underline{p}_{1_1}$, and $\lambda_j = \lambda_1$ into (6.24), we see the following must hold:

$$|\underline{p}_{1_1}, G_1^* \underline{p}_{1_1}, G_1 \underline{v}, G_1^* G_1 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4. \quad (6.55)$$

There exists $\underline{v} \in \mathcal{C}^4$ such that

$$G_1 \underline{v} = \underline{q}_{1_2} + \gamma \underline{q}_{1_1},$$

where γ is chosen such that $G_1 \underline{v} - \text{sp}\{\underline{q}_{1_1}\}$. Using (6.53) we obtain that

$$G_1^* G_1 \underline{v} = \underline{q}_{1_1}.$$

Substituting these expressions for $G_1 \underline{v}$ and $G_1^* G_1 \underline{v}$ into (6.55) and simplifying yields

$$|\underline{p}_{1_1}, G_1^* \underline{p}_{1_1}, \underline{q}_{1_2}, \underline{q}_{1_1}| = 0.$$

This means that there exists constants $\alpha_1, \alpha_2, \alpha_3$, and α_4 , not all zero, such that

$$\alpha_1 \underline{p}_{1_1} + \alpha_2 G_1^* \underline{p}_{1_1} + \alpha_3 \underline{q}_{1_2} + \alpha_4 \underline{q}_{1_1} = \underline{0}. \quad (6.56)$$

If $\alpha_2 = 0$, then $\underline{p}_{1_1}, \underline{q}_{1_2}$, and \underline{q}_{1_1} must be linearly dependent. In this case, if α_1 is also zero, we would have a contradiction since this would imply that \underline{q}_{1_2} and \underline{q}_{1_1} are dependent. So, we assume that $\alpha_1 \neq 0$ and thus

$$\underline{p}_{1_1} = -\left(\frac{\alpha_3}{\alpha_1} \underline{q}_{1_2} + \frac{\alpha_4}{\alpha_1} \underline{q}_{1_1}\right).$$

Notice that both α_3 and α_4 cannot be zero because this would mean that $\underline{p}_{1_1} = \underline{0}$.

We will also assume that $\alpha_3 \neq 0$, since $\underline{p}_{1_1} - \underline{q}_{1_1}$. Substituting this expression for \underline{p}_{1_1} into (6.55) then using (6.53) and simplifying shows that

$$|\underline{q}_{1_2}, \underline{q}_{1_1}, G_1 \underline{v}, G_1^* G_1 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4.$$

Since \underline{q}_{1_1} spans the nullspace of G_1^* , notice that any vector $\underline{v} \in \mathcal{C}^4$ can be written as

$$\underline{w} = G_1 \underline{v} + \gamma \underline{q}_{1_1},$$

for some $\underline{v} \in \mathcal{C}^4$, and some $\gamma \in \mathcal{C}$. Multiplying the second column in the above determinant by γ and adding it to the third column, then substituting

$$G_1^* G_1 \underline{v} = G_1^* (G_1 \underline{v} + \gamma \underline{q}_{1_1})$$

into the above shows that for every $\underline{w} \in \mathcal{C}^4$,

$$|\underline{q}_{1_2}, \underline{q}_{1_1}, \underline{w}, G_1^* \underline{w}| = 0.$$

Next, let $\underline{w} = \underline{q}_{2_2}$ and notice that $G_1^* \underline{q}_{2_2} = \underline{q}_{2_1} + (\bar{\lambda}_2 - \bar{\lambda}_1) \underline{q}_{2_2}$. Substitution into the above and simplifying produces

$$|\underline{q}_{1_2}, \underline{q}_{1_1}, \underline{q}_{2_2}, \underline{q}_{2_1}| = 0.$$

However, this is impossible because these vectors are linearly independent. Therefore, we will assume that $\alpha_2 \neq 0$ in (6.56), which means

$$G_1^* \underline{p}_{1_1} \in \text{sp}\{\underline{p}_{1_1}, \underline{q}_{1_2}, \underline{q}_{1_1}\} \quad (6.57)$$

If we begin by substituting $\underline{p}_j = \underline{p}_{2_1}$ and $\lambda_j = \lambda_2$ into (6.24), an analogous argument shows

$$G_2^* \underline{p}_{2_1} \in \text{sp}\{\underline{p}_{2_1}, \underline{q}_{2_2}, \underline{q}_{2_1}\}. \quad (6.58)$$

Multiplying the third column in (6.55) by $(\bar{\lambda}_1 - \bar{\lambda}_2)I$ and adding it to the fourth column produces

$$|\underline{p}_{1_1}, G_1^* \underline{p}_{1_1}, G_1 \underline{v}, G_2^* G_1 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4. \quad (6.59)$$

Suppose that either:

- (1) \underline{q}_{2_2} and \underline{q}_{2_1} are orthogonal to \underline{q}_{1_1} , or
- (2) \underline{q}_{2_1} is not orthogonal to \underline{q}_{1_1} .

Let

$$G_1 \underline{v} = \underline{q}_{2_2} + \beta \underline{q}_{2_1},$$

where β is chosen so that $G_1 \underline{v}$ is orthogonal to \underline{q}_{1_1} . Substituting this expression for $G_1 \underline{v}$ into (6.59) and simplifying yields

$$|\underline{p}_{1_1}, G_1^* \underline{p}_{1_1}, \underline{q}_{2_2}, \underline{q}_{2_1}| = 0.$$

Since $\underline{p}_{1_1} \in \text{sp}\{G_1^* \underline{p}_{1_1}, \underline{q}_{2_2}, \underline{q}_{2_1}\}$, it follows that

$$G_1^* \underline{p}_{1_1} \in \text{sp}\{\underline{q}_{2_2}, \underline{q}_{2_1}\}.$$

Notice that $\underline{p}_{1_2} \in \text{sp}\{\underline{q}_{2_2}, \underline{q}_{2_1}\}$ which implies that $G_1^* \underline{p}_{1_1} - \underline{p}_{1_2}$, or equivalently,

$$0 = \langle G_1^* \underline{p}_{1_1}, \underline{p}_{1_2} \rangle = \langle \underline{p}_{1_1}, G_1 \underline{p}_{1_2} \rangle = \langle \underline{p}_{1_1}, \underline{p}_{1_1} \rangle.$$

This is a contradiction to our assumption that \underline{p}_{1_1} is an eigenvector of \tilde{R} . Therefore, it follows that \underline{q}_{2_1} must be orthogonal to \underline{q}_{1_1} , and \underline{q}_{2_2} is not orthogonal to \underline{q}_{1_1} . From a symmetric argument, we can also conclude that \underline{q}_{2_1} is orthogonal to \underline{q}_{1_1} , and \underline{q}_{1_2} is not orthogonal to \underline{q}_{2_1} .

Suppose \underline{q}_{2_1} is orthogonal to \underline{q}_{1_1} , but \underline{q}_{2_1} is not orthogonal to \underline{q}_{1_2} , and \underline{q}_{2_2} is not orthogonal to \underline{q}_{1_1} . From (6.24), with $\underline{p}_j = \underline{p}_{2_1}$ and $\lambda_j = \lambda_2$, it follows that

$$|\underline{p}_{2_1}, G_2^* \underline{p}_{2_1}, G_2 \underline{v}, G_2^* G_2 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4.$$

Every $\underline{v} \in \mathcal{C}^4$ can be written as $\underline{v} = \underline{w} + \alpha \underline{z}$, for some $\underline{w}, \underline{z} \in \mathcal{C}^4$ and some $\alpha \in \mathcal{C}$.

Using this expression for \underline{v} , the above can be rewritten as

$$\mathcal{G}(\alpha) = |\underline{p}_{2_1}, G_2^* \underline{p}_{2_1}, G_2(\underline{w} + \alpha \underline{z}), G_2^* G_2(\underline{w} + \alpha \underline{z})| = 0, \quad (6.60)$$

for every $\underline{w}, \underline{z} \in \mathcal{C}^4$, and for every $\alpha \in \mathcal{C}$. Differentiating once with respect to α , then setting $\alpha = 0$ produces

$$\mathcal{G}'(\alpha)|_{\alpha=0} = |\underline{p}_{2_1}, G_2^* \underline{p}_{2_1}, G_2 \underline{z}, G_2^* G_2 \underline{w}| + |\underline{p}_{2_1}, G_2^* \underline{p}_{2_1}, G_2 \underline{w}, G_2^* G_2 \underline{z}| = 0,$$

for every $\underline{w}, \underline{z} \in \mathcal{C}^4$. Since $\underline{q}_{2_1} - \underline{q}_{1_1}$, it follows that \underline{q}_{1_1} is in the range of G_2 . There exists $\underline{w} \in \mathcal{C}^4$ such that $G_2\underline{w} = \underline{q}_{1_1}$. Substituting this into the above yields

$$|\underline{p}_{2_1}, G_2^*\underline{p}_{2_1}, G_2\underline{z}, (\bar{\lambda}_1 - \bar{\lambda}_2)\underline{q}_{1_1}| + |\underline{p}_{2_1}, G_2^*\underline{p}_{2_1}, \underline{q}_{1_1}, G_2^*G_2\underline{z}| = 0,$$

for every $\underline{z} \in \mathcal{C}^4$. This can be rearranged and combined by using the properties of determinants to give

$$|\underline{p}_{2_1}, G_2^*\underline{p}_{2_1}, \underline{q}_{1_1}, G_1^*G_2\underline{z}| = 0, \quad \forall \underline{z} \in \mathcal{C}^4.$$

Since \underline{q}_{2_1} is not orthogonal to \underline{q}_{1_2} , we know that there exists $\underline{z} \in \mathcal{C}^4$ such that

$$G_2\underline{z} = \underline{q}_{1_2} + \gamma\underline{q}_{2_1},$$

where $\gamma \neq 0$, and is chosen so that $G_2\underline{z} \in \text{sp}\{\underline{q}_{2_1}\}$. Substituting this into the above and simplifying yields

$$|\underline{p}_{2_1}, G_2^*\underline{p}_{2_1}, \underline{q}_{1_1}, \underline{q}_{2_1}| = 0.$$

Since \underline{p}_{2_1} , \underline{q}_{1_1} , and \underline{q}_{2_1} are linearly independent, in order for the above vectors to be dependent requires that $G_2^*\underline{p}_{2_1} \in \text{sp}\{\underline{p}_{2_1}, \underline{q}_{1_1}, \underline{q}_{2_1}\}$. Together with (6.58) shows that

$$G_2^*\underline{p}_{2_1} \in \text{sp}\{\underline{q}_{2_1}\}. \quad (6.61)$$

A symmetric argument can be used to obtain $G_1^*\underline{p}_{1_1} \in \text{sp}\{\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{q}_{2_1}\}$.

Together with (6.57) yields

$$G_1^*\underline{p}_{1_1} \in \text{sp}\{\underline{q}_{1_1}\}. \quad (6.62)$$

If either $G_1^*\underline{p}_{1_1} \equiv \underline{0}$, or $G_2^*\underline{p}_{2_1} \equiv \underline{0}$, we obtain a contradiction to our assumption that \underline{q}_{1_1} (\underline{q}_{2_1}) is the only eigenvector of \tilde{R}^* with $\bar{\lambda}_1$ ($\bar{\lambda}_2$). Therefore, we assume that

$$G_2^*\underline{p}_{2_1} = \alpha\underline{q}_{2_1}, \quad \text{and} \quad G_1^*\underline{p}_{1_1} = \beta\underline{q}_{1_1}, \quad (6.63)$$

for some nonzero constants, α and β . Notice from (6.54) that $\langle \underline{p}_{1_1}, G_2^*\underline{p}_{2_1} \rangle = \langle \underline{p}_{1_1}, \alpha\underline{q}_{2_1} \rangle = 0$. Since

$$0 = \langle \underline{p}_{1_1}, G_2^*\underline{p}_{2_1} \rangle = \langle G_2\underline{p}_{1_1}, \underline{p}_{2_1} \rangle = (\bar{\lambda}_1 - \bar{\lambda}_2)\langle \underline{p}_{1_1}, \underline{p}_{2_1} \rangle,$$

and $\lambda_1 \neq \lambda_2$, this implies that $\underline{p}_{1_1} - \underline{p}_{2_1}$.

From (6.54), we see that $\underline{p}_{2_1} - \underline{q}_{1_1}$. Thus, there exists $\underline{v} \in \mathcal{C}^4$ such that

$$G_1 \underline{v} = \underline{p}_{2_1}. \quad (6.64)$$

Substituting (6.64) and (6.63) into (6.59), and simplifying yields

$$|\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{p}_{2_1}, \underline{q}_{2_1}| = 0.$$

This is impossible since the vectors $\{\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{p}_{2_1}, \underline{q}_{2_1}\}$ are an orthogonal set.

Therefore, in any case, the minimal polynomial of \tilde{R} cannot be given by

$$p_4(\lambda) = (\lambda - \lambda_1)^2(\lambda - \lambda_2)^2.$$

Case 4: Suppose the minimal polynomial of \tilde{R} is given by

$$p_4(\lambda) = (\lambda - \lambda_1)^2(\lambda - \lambda_2)(\lambda - \lambda_3).$$

Denote \underline{p}_{1_1} (\underline{p}_{1_2}) as the eigenvector (generalized eigenvector) of \tilde{R} associated with eigenvalue λ_1 , and \underline{p}_{2_1} , and \underline{p}_{3_1} to be the eigenvectors of R with eigenvalues λ_2 , and λ_3 , respectively. Similarly, denote \underline{q}_{1_1} (\underline{q}_{1_2}) to be the eigenvector (generalized eigenvector) of \tilde{R}^* with eigenvalue $\bar{\lambda}_1$, and \underline{q}_{2_1} , and \underline{q}_{3_1} as eigenvectors of \tilde{R}^* with eigenvalues $\bar{\lambda}_2$, and $\bar{\lambda}_3$, respectively.

Recall the following relationships between the eigenvectors and generalized eigenvectors of R and R^* :

$$\begin{aligned} (\tilde{R} - \lambda_1 I)\underline{p}_{1_1} &= \underline{0} & (\tilde{R}^* - \bar{\lambda}_1 I)\underline{q}_{1_1} &= \underline{0} \\ (\tilde{R} - \lambda_1 I)\underline{p}_{1_2} &= \underline{p}_{1_1} & (\tilde{R}^* - \bar{\lambda}_1 I)\underline{q}_{1_2} &= \underline{q}_{1_1} \\ (\tilde{R} - \lambda_2 I)\underline{p}_{2_1} &= \underline{0} & (\tilde{R}^* - \bar{\lambda}_2 I)\underline{q}_{2_1} &= \underline{0} \\ (\tilde{R} - \lambda_3 I)\underline{p}_{3_1} &= \underline{0} & (\tilde{R}^* - \bar{\lambda}_3 I)\underline{q}_{3_1} &= \underline{0} \end{aligned} \quad (6.65)$$

Again, the Jordan decompositions of \tilde{R} , and \tilde{R}^* , show that the eigenvectors (generalized eigenvectors) of \tilde{R} and \tilde{R}^* satisfy the following orthogonality condition:

$$S^{-1}S = \begin{bmatrix} - & \underline{q}_{1_2}^* & - \\ - & \underline{q}_{1_1}^* & - \\ - & \underline{q}_{2_1}^* & - \\ - & \underline{q}_{3_1}^* & - \end{bmatrix} \begin{bmatrix} | & | & | & | \\ \underline{p}_{1_1} & \underline{p}_{1_2} & \underline{p}_{2_1} & \underline{p}_{3_1} \\ | & | & | & | \end{bmatrix} = I. \quad (6.66)$$

Substituting $\underline{p}_j = \underline{p}_{1_1}$, and $\lambda_j = \lambda_1$ in (6.24) shows that the following must hold:

$$|\underline{p}_{1_1}, G_1^* \underline{p}_{1_1}, G_1 \underline{v}, G_1^* G_1 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4. \quad (6.67)$$

Since the range of G_1 is orthogonal to the nullspace of G_1^* , which equals $\text{sp}\{\underline{q}_{1_1}\}$, it follows that there exists $\underline{v} \in \mathcal{C}^4$ such that

$$G_1 \underline{v} = \underline{q}_{1_2} + \gamma \underline{q}_{1_1},$$

where γ is chosen so that $G_1 \underline{v}$ is orthogonal to \underline{q}_{1_1} . Using (6.65) we obtain

$$G_1^* G_1 \underline{v} = \underline{q}_{1_1}.$$

Substituting these expressions for $G_1 \underline{v}$ and $G_1^* G_1 \underline{v}$ into (6.67) and then using the properties of determinants to simplify yields

$$|\underline{p}_{1_1}, G_1^* \underline{p}_{1_1}, \underline{q}_{1_2}, \underline{q}_{1_1}| = 0.$$

This means that there exists constants α_1 , α_2 , α_3 , and α_4 , not all zero, such that

$$\alpha_1 \underline{p}_{1_1} + \alpha_2 G_1^* \underline{p}_{1_1} + \alpha_3 \underline{q}_{1_2} + \alpha_4 \underline{q}_{1_1} = \underline{0}. \quad (6.68)$$

If $\alpha_2 = 0$, it would follow that \underline{p}_{1_1} , \underline{q}_{1_2} , and \underline{q}_{1_1} are dependent. In this case, if $\alpha_1 = 0$, we would have a contradiction since this would imply that \underline{q}_{1_2} and \underline{q}_{1_1} are dependent. Therefore, we assume that $\alpha_1 \neq 0$, and

$$\underline{p}_{1_1} = -\left(\frac{\alpha_3}{\alpha_1} \underline{q}_{1_2} + \frac{\alpha_4}{\alpha_1} \underline{q}_{1_1}\right).$$

Notice that both α_3 and α_4 cannot be zero since this would mean $\underline{p}_{1_1} = \underline{0}$. Also, $\alpha_3 \neq 0$, since $\underline{p}_{1_1} = \underline{q}_{1_1}$. Substituting this expression for \underline{p}_{1_1} into (6.67), then using (6.65) and simplifying produces

$$|\underline{q}_{1_2}, \underline{q}_{1_1}, G_1 \underline{v}, G_1^* G_1 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4.$$

Since \underline{q}_{1_1} spans the nullspace of G_1^* , every $\underline{w} \in \mathcal{C}^4$ can be written as $\underline{w} = G_1 \underline{v} + \gamma \underline{q}_{1_1}$, for some $\underline{v} \in \mathcal{C}^4$ and some $\gamma \in \mathcal{C}$. Multiplying the second column in the above determinant by γ and adding it to the third column, then substituting

$$G_1^* G_1 \underline{v} = G_1^* (G_1 \underline{v} + \gamma \underline{q}_{1_1})$$

into the above shows that for every $\underline{w} \in \mathcal{C}^4$,

$$|\underline{q}_{1_2}, \underline{q}_{1_1}, \underline{w}, G_1^* \underline{w}| = 0.$$

By multiplying the third column by $-(\bar{\lambda}_2 - \bar{\lambda}_1)I$ and adding it to the fourth column, we obtain

$$|\underline{q}_{1_2}, \underline{q}_{1_1}, \underline{w}, G_2^* \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4.$$

Substitute $\underline{w} = \underline{q}_{2_1} + \underline{q}_{3_1}$ into the above, and simplify to produce

$$|\underline{q}_{1_2}, \underline{q}_{1_1}, \underline{q}_{2_1}, \underline{q}_{3_1}| = 0,$$

which is a contradiction since the vectors are linearly independent. Therefore, we will assume that $\alpha_2 \neq 0$ in (6.68), which means

$$G_1^* \underline{p}_{1_1} \in \text{sp}\{\underline{p}_{1_1}, \underline{q}_{1_2}, \underline{q}_{1_1}\}. \quad (6.69)$$

Suppose that either \underline{q}_{2_1} or \underline{q}_{3_1} is not orthogonal to \underline{q}_{1_1} . Without loss of generality, suppose \underline{q}_{2_1} is not orthogonal to \underline{q}_{1_1} . Notice that this implies that \underline{q}_{2_1} is not in the range of G_1 . There exists $\underline{v} \in \mathcal{C}^4$ such that

$$G_1 \underline{v} = \underline{q}_{2_1} + \gamma \underline{q}_{1_1},$$

where $\gamma \neq 0$, and is chosen so that $G_1 \underline{v}$ is orthogonal to \underline{q}_{1_1} . Since $G_1^* \underline{q}_{2_1} = (\bar{\lambda}_2 - \bar{\lambda}_1) \underline{q}_{2_1}$, it follows from (6.65) that

$$G_1^* G_1 \underline{v} = (\bar{\lambda}_2 - \bar{\lambda}_1) \underline{q}_{2_1}.$$

Substitution of these expressions into (6.67) and simplifying produces

$$|\underline{p}_{1_1}, G_1^* \underline{p}_{1_1}, \underline{q}_{1_1}, \underline{q}_{2_1}| = 0.$$

Since $\{\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{q}_{2_1}\}$ is a linearly independent set, in order that the above determinant be zero requires that

$$G_1^* \underline{p}_{1_1} \in \text{sp}\{\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{q}_{2_1}\}. \quad (6.70)$$

If \underline{q}_{2_1} is not orthogonal to \underline{q}_{1_1} , then both (6.69) and (6.70) must hold. Therefore, we must have

$$\begin{aligned} G_1^* \underline{p}_{1_1} &\in \text{sp}\{\underline{p}_{1_1}, \underline{q}_{1_2}, \underline{q}_{1_1}\} \cap \text{sp}\{\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{q}_{2_1}\} \\ &= \text{sp}\{\underline{p}_{1_1}, \underline{q}_{1_1}\}. \end{aligned}$$

Notice that if \underline{q}_{2_1} is orthogonal to \underline{q}_{1_1} but \underline{q}_{3_1} is not orthogonal to \underline{q}_{1_1} , then the above argument could be repeated using \underline{q}_{3_1} instead of \underline{q}_{2_1} . Thus, if either \underline{q}_{2_1} or \underline{q}_{3_1} is not orthogonal to \underline{q}_{1_1} , we can obtain

$$G_1^* \underline{p}_{1_1} \in \text{sp}\{\underline{p}_{1_1}, \underline{q}_{1_1}\}. \quad (6.71)$$

Since $\underline{p}_{1_1} \in \text{sp}\{G_1^* \underline{p}_{1_1}, \underline{q}_{1_1}\}$, it follows that $G_1^* \underline{p}_{1_1} = \alpha \underline{q}_{1_1}$. If $\alpha = 0$, \underline{p}_{1_1} is an eigenvector of \tilde{R}^* with $\bar{\lambda}_1$ which is a contradiction. Therefore, we assume $\alpha \neq 0$, and substitute this expression for $G_1^* \underline{p}_{1_1}$ into (6.67) and then simplify to obtain

$$|\underline{p}_{1_1}, \underline{q}_{1_1}, G_1 \underline{v}, G_1^* G_1 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4.$$

Since \underline{q}_{1_1} spans the nullspace of G_1^* , it follows that any $\underline{w} \in \mathcal{C}^4$ can be written as

$$\underline{w} = G_1 \underline{v} + \gamma \underline{q}_{1_1},$$

for some $\underline{v} \in \mathcal{C}^4$ and $\gamma \in \mathcal{C}$. Multiplying the second column in the above determinant by γ and adding it to the third column, then substituting

$$G_1^* G_1 \underline{v} = G_1^* (G_1 \underline{v} + \gamma \underline{q}_{1_1})$$

into the above produces

$$|\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{w}, G_1^* \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4.$$

Multiplying the third column in the above by $(\bar{\lambda}_1 - \bar{\lambda}_2)I$ and adding it to the fourth column yields

$$|\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{w}, G_2^* \underline{w}| = 0, \quad \forall \underline{w} \in \mathcal{C}^4. \quad (6.72)$$

Let $\underline{w} = \underline{q}_{2_1} + \underline{q}_{3_1}$ in the above to produce

$$|\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{q}_{2_1}, \underline{q}_{3_1}| = 0.$$

However, this is impossible since $\{\underline{p}_{1_1}, \underline{q}_{1_1}, \underline{q}_{2_1}, \underline{q}_{3_1}\}$ are linearly independent. Therefore, \underline{q}_{1_1} must be orthogonal to $\text{sp}\{\underline{q}_{2_1}, \underline{q}_{3_1}\}$.

Suppose that both \underline{q}_{2_1} and \underline{q}_{3_1} are orthogonal to \underline{q}_{1_1} . Multiplying the third column in (6.67) by $(\bar{\lambda}_1 - \bar{\lambda}_2)I$ and adding it to the fourth column produces

$$|\underline{p}_{1_1}, G_1^* \underline{p}_{1_1}, G_1 \underline{v}, G_2^* G_1 \underline{v}| = 0, \quad \forall \underline{v} \in \mathcal{C}^4. \quad (6.73)$$

Since \underline{q}_{2_1} and \underline{q}_{3_1} are both in the range of G_1 , there exists $\underline{v} \in \mathcal{C}^4$ such that

$$G_1 \underline{v} = \underline{q}_{2_1} + \underline{q}_{3_1}.$$

Next, using (6.65), we obtain that

$$G_2^* G_1 \underline{v} = (\bar{\lambda}_3 - \bar{\lambda}_2) \underline{q}_{3_1}.$$

Again, substituting these expressions for $G_1 \underline{v}$ and $G_2^* G_1 \underline{v}$ into (6.73) and using the properties of determinants yields

$$|\underline{p}_{1_1}, G_1^* \underline{p}_{1_1}, \underline{q}_{2_1}, \underline{q}_{3_1}| = 0.$$

Since \underline{p}_{1_1} , \underline{q}_{2_1} , and \underline{q}_{3_1} are linearly independent, this implies that

$$G_1^* \underline{p}_{1_1} \in \text{sp}\{\underline{p}_{1_1}, \underline{q}_{2_1}, \underline{q}_{3_1}\}. \quad (6.74)$$

Notice that (6.69) also holds. Therefore, we obtain

$$\begin{aligned} G_1^* \underline{p}_{1_1} &\in \text{sp}\{\underline{p}_{1_1}, \underline{q}_{1_2}, \underline{q}_{1_1}\} \cap \text{sp}\{\underline{p}_{1_1}, \underline{q}_{2_1}, \underline{q}_{3_1}\} \\ &= \text{sp}\{\underline{p}_{1_1}\}, \end{aligned}$$

which means that $G_1^* \underline{p}_{1_1} = \alpha \underline{p}_{1_1}$. This is impossible for $\alpha \neq 0$ since $\underline{p}_{1_1} \neq G_1^* \underline{p}_{1_1}$. Therefore, $\alpha = 0$, and it follows that \underline{p}_{1_1} is an eigenvector of \tilde{R}^* with $\bar{\lambda}_1$. We have assumed that \underline{q}_{1_1} is the only eigenvector of \tilde{R}^* with $\bar{\lambda}_1$. Since $\underline{p}_{1_1} \neq \underline{q}_{1_1}$, this is a contradiction.

In all cases, if \tilde{R} has a single nonlinear elementary divisor of size two, we have obtained a contradiction. Therefore, the minimal polynomial of \tilde{R} is not given by

$$p_4(\lambda) = (\lambda - \lambda_1)^2(\lambda - \lambda_2)(\lambda - \lambda_3).$$

The above arguments show that \tilde{R} cannot have any nonlinear elementary divisors. Since \tilde{R} is the matrix representation of \tilde{A} , it follows that \tilde{A} cannot have any nonlinear elementary divisors. Therefore, \tilde{A} is diagonalizable.

Suppose A has a nonlinear elementary divisor. This means that there exists, \underline{z}_{j_1} , and \underline{z}_{j_2} , and some λ_j for which

$$(A - \lambda_j I) \underline{z}_{j_2} = \underline{z}_{j_1}, \quad \text{and}$$

$$(A - \lambda_j I) \underline{z}_{j_1} = \underline{0}.$$

The above argument showed that for every \underline{p}_0 such that $d(\underline{p}_0, A) = 4$, the restriction of A to

$$\mathcal{V}_4 = \text{sp}\{\underline{p}_0, A\underline{p}_0, A^2\underline{p}_0, A^3\underline{p}_0\},$$

is diagonalizable. Denote $\{\underline{p}_1, \underline{p}_2, \underline{p}_3, \underline{p}_4\}$ as the eigenvectors that span \mathcal{V}_4 . Let

$$\hat{\mathcal{V}}_4 = \text{sp}\{\underline{p}_1, \underline{p}_2, \underline{z}_{j_1}, \underline{z}_{j_2}\},$$

and let \tilde{A} be the restriction of A to $\hat{\mathcal{V}}_4$. $\hat{\mathcal{V}}_4$ is a 4-dimensional invariant subspace of both \tilde{A} and \tilde{A}^\dagger . The above argument can be used to show that \tilde{A} has no nonlinear elementary divisors. This is a contradiction to the assumption that A has a nonlinear elementary divisor. Therefore, A has a complete set of eigenvectors, thus, A is diagonalizable. \square

COROLLARY 6.9 $A \in \mathcal{CG}_{\text{MR}}(b, t)$, with $(b, t) \leq (1, 1)$, if and only if

- (1) $d(A) \leq 3$, or
- (2) A is either B -normal(0, 1) or B -normal(1, 1).

Proof: Sufficient conditions follow from Theorem 6.3. From the first statement in Theorem 6.3, restricting $(b, t) \leq (1, 1)$, we obtain B -normal(0, 0), B -normal(0, 1), and B -normal(1, 1) matrices. Notice that if A is B -normal(0, 0), equivalently, B -normal(0), then A^\dagger can be written as a constant polynomial. This implies that

$$A = \alpha I, \quad \text{for some } \alpha \in \mathcal{C},$$

and thus, A has only 1 distinct eigenvalue. A close look at the second statement in Theorem 6.3 and Corollary 4.4, shows that rank 1 perturbations of B -normal(0) matrices are also contained in the class $\mathcal{CG}_{\text{MR}}(b, t)$, with $(b, t) = (1, 1)$. Since B -normal(0) matrices have only one distinct eigenvalue, any rank 1 perturbation of a B -normal(0) matrix, can have at most 2 distinct eigenvalues, thus, $d(A) \leq 2$.

Notice that if $d(A) \leq 3$, then for any \underline{p}_0 , $d = d(\underline{p}_0, A) \leq 3$. This means

$$\mathcal{K}_d(\underline{p}_0, A) = \text{sp}\{\underline{p}_0, A\underline{p}_0, A^2\underline{p}_0\} = \text{sp}\{\underline{p}_0, \underline{p}_1, \underline{p}_2\},$$

and the recursion $MR(1, 1)$ will yield a B -orthogonal basis for any \underline{p}_0 in 2 or less steps.

Suppose $A \in \mathcal{CG}_{\text{MR}}(b, t)$, with $(b, t) \leq (1, 1)$. Corollary 6.6 shows that either $d(A) \leq 3$, or $A \in \mathcal{CG}_{\text{SR}}(s, t)$, with $s = b$ and $t = 1$. From Theorem 6.8, it follows that $d(A) \leq 3$, or A is B -normal($b, 1$). \square

We note here that the classes $\mathcal{CG}_{\text{MR}}(b, t)$ with $(b, t) \leq (1, 1)$, and $\mathcal{CG}_{\text{SR}}(s, t)$ with $(s, t) \leq (1, 1)$, are not equivalent. B -normal(1) matrices are included in $\mathcal{CG}_{\text{SR}}(s, t)$ with $(s, t) = (1, 0)$. However, from Corollary 6.9, it follows that B -normal(1) matrices are not in $\mathcal{CG}_{\text{MR}}(b, t)$ with $(b, t) \leq (1, 1)$. From Theorem 6.3 we see that these matrices are included in $\mathcal{CG}_{\text{MR}}(2, 0)$.

7. Conclusion

The conjugate gradient method is implemented via the construction of a B -orthogonal basis for the underlying Krylov subspace. In this thesis, we have considered when this construction can be accomplished using some form of a short recurrence, yielding an economical conjugate gradient algorithm.

The theory from Faber and Manteuffel [5] applies to a special form of recursion, called an $(s + 2)$ -term recursion (2.11). This theory shows that a practical implementation of a conjugate gradient method using a single $(s + 2)$ -term recursion, is limited to a very small class of matrices, called B -normal(s) matrices [5]. The work done by Gragg, Jagels and Reichel ([10], [16], [17]) on unitary and shifted unitary matrices, demonstrated that the single $(s + 2)$ -term recursion was not general enough to include all possible forms of short recurrences. For these matrices, a short double recursion can be used to construct an orthogonal basis, whereas, it is not possible to do this with a short $(s + 2)$ -term recursion. This work motivated the formulation in this thesis of a special type of multiple recursion, $MR(b, t)$ (6.4). Using this form, we demonstrated that the class of matrices for which an efficient conjugate gradient algorithm is known to exist, can be extended.

Sufficient conditions on the system matrix A were determined in order that a short multiple recursion $MR(b, t)$ will yield a B -orthogonal basis. These conditions are for the matrix A to be either B -normal(ℓ, m), or a generalization of a B -normal(ℓ, m) matrix (see Sections 4.3, and 4.4). This includes the B -normal(s) matrices characterized in [5], as well as unitary and shifted unitary matrices studied in ([10], [16], and [17]). In addition, this class includes low rank perturbations of

B -normal(ℓ, m) matrices. An example of this type is a low rank perturbation of a self-adjoint matrix.

To determine if there are any other matrices for which a B -orthogonal basis can be constructed using a recursion of this type, we must determine if the sufficient conditions are also necessary. This question was answered for only a restricted subset of the multiple recursions $MR(b, t)$, in particular, necessary conditions were determined only for $(b, t) \leq (1, 1)$. Further research is needed to complete the analysis of necessary conditions when $b, t > 1$.

This research opens the door to the possibility that short recurrences exist for even a wider class of matrices. The multiple recursions $MR(b, t)$ studied in this thesis, involve one recursion for the direction vector \underline{p}_{j+1} at each step, and t recursions for auxiliary vectors, \underline{q}_{j+1_i} , $i = 0, \dots, t-1$. The recursions for the auxiliary vectors each have only two terms. We might consider recursions for the \underline{q}_{j+1_i} 's that utilize more terms. An example of such a recursion is given in Chapter 2, (2.12). The recursions $MR(b, t)$, are actually a special case of this type. Future research on alternate forms of short recursions might begin with the study of this more general form of short multiple recursion.

A. Properties Of Determinants

The following is a list of properties of determinants that will be used throughout the main proofs in this chapter. These properties can be verified by checking a basic linear algebra text (cf. [22], pp. 158-174).

Suppose A is a $p \times p$ matrix, and denote, $|A|$ as the determinant of A .

- (1) If A has a row (column) of zeros, then $|A| = 0$.
- (2) If A has two identical rows (columns), then $|A| = 0$.
- (3) For any number β , $|\beta A| = \beta^p |A|$.
- (4) $|A| = |A^T|$.
- (5) If D is obtained from A by interchanging two rows (columns) of A , then $|D| = (-1)|A|$.
- (6) If D is obtained from A by replacing one row (column) of A by a number β times that row (column), then $|D| = \beta|A|$.
- (7) If D is obtained from A by replacing one row (column) of A by that row (column) plus some multiple of a different row (column), then $|D| = |A|$.
- (8) If A and D are equal except for possibly the entries in the j 'th row (column), and if C is defined as the matrix identical to A and D except that its j 'th row (column) is the sum of the j 'th rows (columns) of A and D , then $|C| = |A| + |D|$.
- (9) $|AD| = |A||D|$.

(10) $(\partial/\partial t) |A(t)|$ is the sum of the p determinants:

$$\begin{vmatrix} a'_{1,1}(t) & a_{1,2}(t) & \cdots & a_{1,p}(t) \\ \vdots & \vdots & & \vdots \\ a'_{p,1}(t) & a_{p,2}(t) & \cdots & a_{p,p}(t) \end{vmatrix} + \cdots + \begin{vmatrix} a_{1,1}(t) & \cdots & a_{1,p-1}(t) & a'_{1,p}(t) \\ \vdots & & \vdots & \vdots \\ a_{p,1}(t) & \cdots & a_{p,p-1}(t) & a'_{p,p}(t) \end{vmatrix}.$$

REFERENCES

- [1] 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.
- [2] T. Barth and T. A. Manteuffel, **Variable Metric Conjugate Gradient Methods**, Proceedings for the conference in Matrix Analysis and Parallel Computing, Advances in Numerical Methods for Large Sparse Sets of Linear Equations, No. 10, Keio University, Japan, (March 1994).
- [3] L. Elsner and Kh. D. Ikramov, **On a Condensed Form for Normal Matrices under Finite Sequences of Unitary Similarities**, University of Bielefeld, West Germany, Preprint 94-077.
- [4] L. Elsner and Kh. D. Ikramov, **On Nonnormal Matrices that can be Reduced to Band Form under Finite Sequences of Elementary Unitary Similarities**, University of Bielefeld, West Germany, Preprint 96-008.
- [5] 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, (1984) pp. 352-362.
- [6] V. Faber and T. A. Manteuffel, **Orthogonal Error Methods**, SIAM J. Numer. Analysis, Vol. 24, No. 1, (1987), pp. 170-187.
- [7] R. W. Freund and N. M. Nachtigal, **QMR: a quasi-minimal residual method for non-Hermitian linear systems**, Numer. Math., 60, (1991), pp. 315-339.
- [8] R. W. Freund, M. H. Gutknecht, and N. M. Nachtigal, **An Implementation of the Look-Ahead Lanczos Algorithm for Non-Hermitian Matrices**, SIAM J. Sci. Comput., Vol. 14, No. 1, (January 1993), pp. 137-158.
- [9] G. H. Golub and C. F. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, Maryland, (1989).
- [10] W. B. Gragg, **Positive definite Toeplitz matrices, the Arnoldi process for isometric operators, and Gaussian quadrature on the unit circle**, J. Comp. Appl. Math, 46 (1993), pp. 183-198.

- [11] J. F. Grcar, **Operator Coefficient Methods for Linear Equations**, Sandia National Laboratory Report SAND89-8691, (November 1989).
- [12] A. Greenbaum, **Behavior of Slightly Perturbed Lanczos and Conjugate-Gradient Recurrences**, Linear Algebra and Its Applications, 113 (1989), pp. 7-63.
- [13] M. H. Gutknecht, **Stationary and Almost Stationary Iterative (k, l) -Step Methods for Linear and Nonlinear Systems of Equations**, Numer. Math. 56 (1989), pp. 179-213.
- [14] M. R. Hestenes and E. L. Stiefel, **Methods of Conjugate Gradients for Solving Linear Systems**, J. Res. Nat. Bur. Standards, Vol. 49, (1952), pp. 409-436.
- [15] A. S. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell Publishing Company, New York, (1964).
- [16] C. F. Jagels and L. Reichel, **The isometric Arnoldi process and an application to iterative solution of large linear systems**, in Iterative Methods in Linear Algebra, eds. R. Beauwens and P. de Groen, Elsevier, Amsterdam, (1992), pp. 361-369.
- [17] C. F. Jagels and L. Reichel, **A Fast Minimal Residual Algorithm For Shifted Unitary Matrices**, Numer. Linear Algebra Appl., 1 (1994), pp. 555-570.
- [18] W. D. Joubert and D. M. Young, **Necessary and sufficient conditions for the simplification of generalized conjugate-gradient algorithms**, Linear Algebra Appl., (1988), pp. 449-485.
- [19] W. D. Joubert, **Generalized conjugate gradient and Lanczos method for the solution of nonsymmetric systems of linear equations**, Ph.D. thesis and Report CNA-238, Center for Numerical Analysis, University of Texas, Austin, TX, (January 1990).
- [20] W. D. Joubert, **Lanczos Methods for the Solution of Nonsymmetric Systems of Linear Equations**, SIAM J. on Matrix Analysis and Appl., vol. 13, no. 3, (July 1992), pp. 926-943.
- [21] G. D. Mostow, J. H. Sampson, and Jean-Pierre Meyer, Fundamental Structures of Algebra, McGraw-Hill, New York, (1963).
- [22] B. Noble and J. W. Daniel, Applied Linear Algebra, Prentice-Hall, New Jersey, (1968).

- [23] Y. Saad and M. H. Schultz, **GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems**, SIAM J. Sci. Statist. Comput., (1986), No. 7, pp. 856-869.
- [24] D. S. Watkins, **Some Perspectives On The Eigenvalue Problem**, SIAM Reviews, (Sept. 1993), Vol. 35, No. 3, pp. 430-471.