Generalized and Quadratic Eigenvalue Problems with Hermitian Matrices

Ali Mohamud Farah

School of Mathematics and Statistics

The University of Birmingham

Birmingham B15 2TT

U.K.

February 21, 2012

CONTENTS

1. INTRODUCTION

Eigenvalue and eigenvector computations are extremely important and have various applications in engineering, physics and other scientific disciplines. What makes eigenvalue computation so interesting is its versatile nature of applicability to new phenomena. The importance of eigenvalue computation has attracted significant research interest from many applied mathematicians and numerical analysts who are tirelessly devising new algorithms and computer programs to satisfy the ever increasing demand for powerful software tools for scientists and engineers. Since a thorough treatment of the subject of eigenvalue computation is beyond my intention, I will concentrate on surveying some of the most prominent methods used for calculations of the eigenvalues of definite Hermitian and symmetric matrices as they relate to the generalized eigenvalue problem (GEP) and the quadratic eigenvalue problem (QEP).

The thesis is divided into six chapters which are closely linked. The first chapter provides an introduction, the aim of the study and the significance of the study. The primary objective of chapter two is a literature review. The vast majority of the material in this chapter is a review of the modern literature on the standard eigenvalue problem (SEP). Definitions of the important terms and some of the theorems that underpin the theory of standard eigenvalue problem are presented in section one. The properties of Hermitian and symmetric matrices are discussed in Section 2.2.6 including the fact that the eigenvalues of

Hermitian matrices are necessarily real and can be arranged in any prescribed order. The fact that the linear combination of Hermitian matrices is always Hermitian is also presented in this section. Furthermore, the important result that Hermitian matrices automatically acquire all the properties of normal matrices is explored.

An important applications of Hermitian matrices discussed in section 2.2.6 (which we will also encounter in Chapters 4, 5) is the fact that a general matrix can be written as sum of two Hermitian matrices. This allows us to project the numerical range of the matrix on to the open right half plain ($ORHP$). Theorem 3 shows that a matrix $A \in M_n$ can be written as $H(A) + iS(A)$. $H(A)$ is Hermitian and corresponds to the real part of matrix A; whereas, $S(A)$ is skew-Hermitian and corresponds to the imaginary part of A. In Section 2.3.3 we deduce the importance of writing a Hermitian matrix as a sum of a Hermitian and skew-Hermitian matrices which is essentially to address the interesting and extremely useful topic of perturbation theory of Hermitian matrices; specially the relationship between the eigenvalues of a sum of two Hermitian matrices. Furthermore, we address in this chapter the important theorems: Weyl's perturbation theorem, the interlacing theorem, the Rayleigh-Ritz and Courant-Fisher minimax theorem.

The spectral theorem of Hermitian matrices is addressed in Section 2.3.1. We discuss some of the prominent techniques used for computing eigenvalues and eigenvectors of Hermitian and real symmetric matrices. We investigate the methods used to solve the problems involving the eigenvalues and eigenvectors of Hermitian Matrices including matrix factorization that can be employed to simplify matrices.

Our principal aim is to investigate the different methods employed for transforming matrices into simpler canonical forms, particularly the matrix transformations that preserve the eigenvalues. This topic is discussed in section 3 where we address the matrix transformation techniques that simplify the computation of the eigensystems.

Chapter 3 is concerned with the techniques of expressing matrix decomposition as it relates to the solution of systems of linear equations. Some of the most frequently used techniques are treated including block LU factorization, Cholesky factorization and QR factorization. These decompositions are applicable for nonsigular and positive definite matrices respectively. Section 3.2 treats the decompositions used for general rectangular matrices such as singular value decomposition (SVD) and the polar decomposition (PD).

Chapter 4 talks about the field of values also known as numerical range. We address this with intent to establish the regions in the complex plane where the eigenvalues of general matrices are found. We pay special attention to the field of values of Hermitian matrices. In section 4.3 we discuss the key idea of simultaneous diagonalization by congruence which establishes the link between the field of values and the solution of the Hermitian form x^*Ax. We assert that if the two matrices concerned are not positive definite then a linear combination can be found which transforms the pair into positive definite. In section 4.4 we discuss the methods for calculating numerical radius, crawford number and numerical range using MATLAB syntaxes.

Chapter 5 is dedicated to the Generalized Eigenvalue Problem (GEP). The essence of the chapter is concerned with the techniques used in computing the eigenvalues and eigenvectors of Hermitian pencils. Special emphasis is put on the (GEP) with at least one of the matrices is Hermitian positive definite. MATLAB based computer programs are provided to calculate numerical radius and Crawford number. In the quest of finding powerful techniques that compute the eigenvalues of GEP of definite Hermitian matrices the congruence transformation is so far the best method. Positive definiteness of the pair is achieved if and only if the scalar quantity $c(A, B)$ known as Crawford number is positive. The strategy is to test for definiteness, if the pair is positive definite then congruence diagonalization is used to transform the GEP into SEP which can be solved using the methods discussed in Chapter 2. Section 5.3 provides an insight into the different methods used for solving GEP including the QZ method, congruence transformation and with positive definite matrices.

Chapter 6 tackles the important concept of the Quadratic Eigenvalue Problem (QEP). linearization techniques are treated in section 6.2. The linearization method transforms the $2n$ QEP into $2n \times 2n$ GEP. The advantage of this approach is that the techniques developed in the SEP and the GEP can be employed. The different types of linearizations are reviewed while focusing on the linearization for hyperbolic and elliptic quadratic eigenvalue problem $(HQEP)$. Moreover, factorization technique is explored in section 6.2.1 which is suitable for the solution of hyperbolic quadratic eigenvalue problem. hyperbolic quadratic eigenvalue problem is treated in section 6.3. The discussion includes testing for hyperbolicity and the overdamped hyperbolic quadratic eigenvalue problem.

Chapter 7 presents conclusion and discussion of the main findings of the thesis.

Throughout this thesis roman capital letters are used for matrices (A, B, etc), roman lowercase letters denote vectors (u, v, x, etc) and lowercase Greek letters represent scalars $(\alpha, \beta, \gamma, etc)$.

1.1 The Purpose of the Study

The aim of this study is partly to examine the existing methods used to solve the generalized eigenvalue problem GEP and the Quadratic eigenvalue problem QEP with definite Hermitian matrices. Furthermore, the research investigates new algorithms and computer programs that reduce the cost of eigenpair (eigenvalue, eigenvector) computations and improve the speed of eigenpair solutions.

1.2 Notations

$\mathbb{R}$	the real numbers
$\mathbb{R}^n$	vector space over real numbers
$\mathbb{C}$	the complex numbers
$\mathbb{C}^n$	vector space over $\mathbb{C}$
$M_{m,n}$	rectangular m-by-n complex matrices
M_n	square n-by-n complex matrices
I_n	identity matrix with dimension n
A^T	transpose of A. If $A = [a_{ij}]$, then $A^T = [a_{ji}]$
A^*	complex conjugate transpose of A i.e. $\bar{A}^T$
A^{-1}	inverse of A
$\|A\|$	entrywise absolute value of $\|a_{ij}\|$
e	vector of all ones
∂F	boundary of the set F
$Co(F)$	convex hull of the set F
$det(A)$	determinant of A
$diag(A)$	diagonal matrix A
$F(A)$	field of values of $A \in M_n$
$H(A)$	Hermitian part of $A \in M_n$
$S(A)$	skew-Hermitian part of $A \in M_n$
$\|\| \cdot \|\|_2$	l_2 (Euclidean) norm of C^n
$\|\| \cdot \|\|_1$	l_1 norm on $\mathbb{C}^n$ or $M_{m,n}$
$\|\| \cdot \|\|_\infty$	l_∞ norm on $\mathbb{C}$ or $M_{m,n}$
$rankA$	rank of $A \in M_{m,n}$
$r(A)$	numerical radius of $A \in M_n$

$\sigma(A)$	spectrum of $A \in M_n$
$\rho(A)$	spectral radius of $A \in M_n$
trA	trace of $A \in M_n$

2. STANDARD EIGENVALUE PROBLEM (SEP)

2.1 Introduction

Section 2.2.1 presents definitions and theorems that are indispensable for the forthcoming chapters. A detailed discussion of the underlying theory and the computational techniques are raised and thoroughly discussed. The definitions of important terms are invoked such as Hermitian matrix, unitary matrix, positive definite matrix, normal matrix to mention a few.

Since Hermitian and symmetric matrices are the central theme of the study, their differences and similarities are discussed in Section 2.2.6. The Hermitian matrices have entries from the complex field, whereas the symmetric matrices are real. It is possible that a matrix with real elements have complex eigenvalues. Therefore, the discussion is concentrated on the matrices with complex elements.

Section 3 is geared towards the triangular factorizations of matrices. Since Hermitian and real symmetric matrices exhibit important properties, more emphasis is put on these types of matrices.

2.2 Mathematical Preliminaries

Hermitian and real symmetric matrices form a very important class of matrices that have many practical applications. Hermitian matrices are special type of normal matrices. The property of normal matrices is that they are closed under unitary equivalence. Furthermore, normal matrices generalize symmetric and Hermitian matrices. The other types of matrices that are normal include: skew Hermitian matrices, unitary matrices, symmetric matrices, skew symmetric matrices and orthogonal matrices. The eigenvalues of a Hermitian matrix can be seen as the solution of a constrained optimization problem.

2.2.1 Definitions of Important Terms and Some Theorems

Definition 1 (eigenvector, eigenvalue)**.** *Let $A \in M_n$ be square matrix and let $x \in \mathbb{C}^n, x \neq 0$, be a vector with elements in the complex field. Furthermore, suppose that the equation $Ax = \lambda x$ has a solution. We say λ is an eigenvalue of A and x an eigenvector of A associated with λ. Moreover, the set that constitute the eigenvalues of A is called the spectrum of A and is denoted by $\sigma(A)$.*

Definition 2 (rank of a matrix)**.** *The rank of $A \in M_n$ is defined as the largest number of linearly independent columns of A. Equivalently, the rank of $A \in M_n$ is the largest number of linearly independent rows of A. Therefore, rank $A =$ rank A^T.*

Definition 3 (trace of a matrix)**.** *Let $A \in M_n$, the trace of A written as trA or trace A is the sum of the elements of the main diagonal of A, i.e., $trA = \sum_{i=1}^{n} a_{ii}$.*

Definition 4 (characteristic polynomial)**.** *Let $A \in M_n$, then the polynomial $det(A - \lambda I) = 0$ is called characteristic polynomial of A and is denoted as $\chi(A)$.*

Definition 5 (hermitian matrix)**.** *A matrix $A \in M_n$ is said to be Hermitian if $A = A^*$, where A^* denotes the complex conjugate transpose of A $(a_{ij} = \bar{a}_{ji}, i \neq j)$. A matrix $A \in M_n$ is said to be skew Hermitian if $A = -A^*$.*

Definition 6 (symmetric matrix). *A matrix $A \in M_n$ is symmetric if $A = A^T$, in other words $a_{ij} = a_{ji} \; \forall \; i \neq j$.*

A real symmetric matrix is necessarily Hermitian, thus conforms with all the properties of Hermitian matrices.

Definition 7 (upper triangular matrix). *A matrix $A \in M_n$ is said to be upper triangular if $a_{ij} = 0$ whenever $j < i$, and lower triangular if $a_{ij} = 0$ whenever $j > i$.*

Definition 8 (diagonal matrix). *A matrix $A \in M_n$ is said to be diagonal if $a_{ij} = 0$ whenever $i \neq j$. This means that the matrix is both upper and lower triangular.*

Definition 9 (unitary matrix). *A matrix $U \in M_n$ is said to be unitary if $U^*U = I$.*

It follows from above that unitary matrices have the extremely important property that the inverse of a unitary matrix is equal to its conjugate transpose see [20] pages 66-68.

Definition 10 (unitary equivalence). *Let $A, B \in M_n$, we say A is unitarily equivalent to B if there exist a unitary matrix U such that $A = U^*BU$.*

Definition 11 (normal matrix). *A matrix $A \in M_n$ is said to be normal if*

$$A^*A = AA^*.$$

Proposition 1. *Normal matrices have the following properties:*

1. *Normal matrices are unitarily similar to a diagonal matrix.*

2. *There is an orthonormal set of n eigenvectors of A.*

3. *$A \in M_n$ is normal if and only if every matrix that is unitarily equivalent to A is normal.*

4. *The sum or product of two normal matrices are not necessarily normal, however, when the two matrices are normal and also commute i.e. $A^*B = B^*A$, then $A + B$ and A^*B are also normal.*

Proof: See [20, Sect.2.5]. □

Example 1. *These are some examples of normal matrices which are of special interest to us:*

1. *All unitary matrices are normal since $U^*U = UU^* = I$.*

2. *All Hermitian matrices are normal since $A^* = A$ directly implies that $A^*A = AA^*$.*

3. *All skew-Hermitian matrices are normal since $A^* = -A$ implies that $A^*A = -A^2 = AA^*$.*

2.2.2 Vector and Matrix Norms

Norms are essential for computing the size of a vector or the distance on the space of a matrix see [20] page 257, [12] page 53-54. They are often used for estimating errors. For example if one cannot find the exact solution for the optimization problem $\min Ax = b$ and tries to approximate, hence finds $\hat{x}$ such that $A\hat{x} \approx b$. It is important to measure how this approximates the true solution. Moreover, if the matrix concerned is nearly singular this may result a poor answer. Among other things matrix norms are essential for computing functions of matrices, matrix powers and they are necessary for the study of higher order matrix polynomials including generalized and quadratic eigenvalue problems. Furthermore, norms of matrices are used as bounds for the field of values of matrices, and for the spectrum of a matrix which we will encounter in the later chapters.

Norms can be seen as functions $\| \cdot \| : A \to \mathbb{R}$ that satisfy the following relations.

The following are matrix norms which will be needed for forthcoming chapters (among other things for the solution of field of values, generalized Hermitian eigenvalue problem and quadratic eigenvalue problem).

1. The l_1 norm for $A \in M_n$ which is the maximum absolute value of the column sum
$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|,$$

2. The l_∞ for $A \in M_n$, which is the maximum absolute value of the row sum
$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{i=1}^{n} |a_{ij}|,$$
This implies that $||A||_1 = ||A^*||_\infty$

3. The Euclidean or l_2 norm for $A \in M_n$
$$\|A\|_2 = \sqrt{\sum_{i,j=1}^{n} |a_{i,j}|^2} \quad \Leftrightarrow \quad \sqrt{\lambda_{max}(A^*A)}.$$

4. Frobenius norm for $A \in M_n$

$$\|A\|_F = \left(\sqrt{\sum_{ij=1}^{n} |a_{ij}|^2} \right)^{\frac{1}{2}} .$$

Matrix norms are equivalent as follows:

$$||A||_2 \leq ||A||_F \leq \sqrt{n}||A||_2$$

$$\frac{1}{\sqrt{n}}||A||_1 \leq ||A||_2 \leq \sqrt{n}||A||_1$$

$$\frac{1}{\sqrt{n}}||A||_\infty \leq ||A||_2 \leq \sqrt{n}||A||_\infty .$$

For more detailed discussion on matrix norms see [43] pages 55-61; [20] pages 257-342;

2.2.3 Similarity Transformation

Definition 12. *A matrix $A \in M_n$ is said to be similar to a matrix $B \in M_n$ if there exist a nonsingular matrix $S \in M_n$ called a similarity transformation such that*

$$SA = BS.$$

Similar matrices represent the same linear transformation but with respect to a different basis. If the similarity matrix S is unitary then we say that A, B are unitarily equivalent. Unitary equivalence will be discussed later. Similar matrices share many nice properties. For example, if two matrices A and B are similar, then they have the same eigenvalues, rank, trace and determinant see [39] pages 235-236. However, having the same eigenvalues is a necessary but not sufficient condition as matrices can have the same eigenvalues but can still be non-similar.

Proposition 2. *Similarity transformation is an equivalence relationship.*

- *A is similar to itself* (reflexivity).
- *If A is similar to B, then B is similar to A* (symmetry).
- *if A is similar to B and B is similar to C then A is similar to C* (transitivity).

Theorem 1. *Let $A, B \in M_n$. If B is similar to A, then the characteristic polynomial of B is the same as that of A.*

Proof: B is similar to A means $B = P^{-1}AP$, then the characteristic polynomial of B is

$$\begin{aligned} \chi(B) &= det(B - \lambda I) = det(P^{-1}AP - \lambda P^{-1}PI) = detP^{-1}(A - \lambda I)P \\ &= detP^{-1}det(A - \lambda I)detP = det(P^{-1})det(P)det(A - \lambda I) \\ &= det(I)det(A - \lambda I) = det(A - \lambda I) = \chi(A)\,. \quad \Box \end{aligned}$$

The similarity transformation is an important technique which is frequently used for extracting eigenvalues and other important properties from matrices. One way of doing this is to determine similarity transformation that finds a similar matrix of special form (that is either upper triangular or diagonal). Once found the similar matrix of special form, the calculation of eigenvalues, determinant and rank of the matrix becomes trivial. The eigenvalues of a triangular matrix equate to its diagonal entries; moreover, its determinant corresponds to the product of the diagonal entries. The rank of a triangular matrix is the number of non zero elements on the diagonal. The trace of a triangular matrix equates the sum of its diagonal entries.

Definition 13. *Let $U \in M_n$ be a unitary matrix, and let $A, B \in M_n$, then A is said to be unitarily similar to B if $A = U^*BU$.*

Unitary similarity (or unitary equivalence) is a finer part of similarity transformation. The change of bases with regard to unitary equivalence implies changes from one orthonormal bases to another. The unitary similarity is preferable compared to general similarity for the following reasons.

- Unitary similarity preserves Hermitian property. If $A \in M_n$ is Hermitian then UAU^* is also Hermitian whereas SAS^{-1} (with a nonsingular $S \in M_n$) may or may not be Hermitian.
- It is easier to compute unitary similarity than general similarity, since the inverse of a unitary matrix is simply its complex conjugate transpose.

Normal matrices are frequently used as a test for diagonalizability. If a similarity matrix is a normal matrix such as Hermitian then its similarity transformation yields a diagonal matrix.

2.2.4 Schur's Unitary Triangularization

The Schur triangularization process provides us with a means of finding the eigenvalues of a matrix by transforming it into a simpler form such as upper triangular form. This part is primarily concerned with the Schur form, which says that given any matrix $A \in M_n$ is unitarily equivalent to a triangular matrix T where the diagonal entries of T are the eigenvalues of A. Moreover, if we take

$A \in \mathbb{R}$ with real eigenvalues, then to achieve the triangularization we can choose a real and orthogonal matrix U. The following is Schur's theorem :

Theorem 2 (Schur triangularization)**.** *Let $A \in M_n$ has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ in any prescribed order. Then exist a unitary matrix $U \in M_n$ and upper triangular T such that*

$$U^* A U = T,$$

where T is upper triangular and $t_{ii} = \lambda_i$ are the eigenvalues of A.

Proof: Let x_1 be a unit eigenvector of A corresponding to the eigenvalue λ_1, i.e, $Ax_1 = \lambda_1 x_1$. If $\|x_1\| \neq 1$ normalize x_1 by setting $z = \frac{x_1}{\|x_1\|}$.

Let Q_1 be unitary matrix with first column x_1. Then $Q_1^* A Q_1 = Q_1^* A [x_1, \ldots, x_n]$ and so

$$\begin{aligned} Q_1^*[Ax_1, Ax_2 \ldots Ax_n] &= Q_1^*[\lambda_1 x_1, Ax_2, \ldots, Ax_n] \\ &= [x_i^* \lambda_1 x_1, x_i^* A x_2, \ldots, x_i^* A x_n] \,. \end{aligned}$$

Thus

$$Q_1^* A Q_1 = \begin{pmatrix} \lambda_1 & * & \ldots & * \\ 0 & & & \\ \vdots & (A_1) & & \\ 0 & & & \end{pmatrix}.$$

Next step is to transform the (A_1) square matrix repeating the previous procedure. The matrix (A_1) has eigenvalues $\lambda_2, \ldots, \lambda_n$. This determines $Q_2 \in M_{n-1}$

which is unitary such that

$$Q_2^* A_1 Q_2 = \begin{pmatrix} \lambda_2 & * \\ 0 & (A_2) \end{pmatrix}.$$

Now let

$$\tilde{Q}_2 = \begin{pmatrix} 1 & 0 \\ 0 & Q_2 \end{pmatrix}.$$

Continuing this procedure we get the desired upper triangular form with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ on the diagonal.

$$\tilde{Q}_{n-1}^* \ldots \tilde{Q}_3^* \tilde{Q}_2^* \tilde{Q}_1^* A \tilde{Q}_1 \tilde{Q}_2 \tilde{Q}_3 \ldots \tilde{Q}_{n-1} = U^* A U$$

where $U = Q_1 \tilde{Q}_2 \tilde{Q}_3 \ldots \tilde{Q}_{n-1}$ is unitary matrix since the product of unitary matrices is unitary. □

Theorem 3. *If $A \in M_n$ is normal then it is unitarily diagonalizable.*

Proof: Let $T = U^* A U$ be a Schur decomposition of A with an upper triangular matrix $T = [t_{ij}] \in M_n$. Recall that any matrix that is unitarily similar to a normal matrix is necessarily normal. Therefore, T is also normal, i.e., $T^* T = T T^*$. We will show that a triangular normal matrix must be diagonal. Equating the $k-$th diagonal entries of $T^* T$ and $T T^*$, we get

$$\overline{t}_{kk} t_{kk} = t_{kk} \overline{t}_{kk} + \sum_{j=k+1}^{n} t_{kj} \overline{t}_{kj} = |t_{kk}|^2 + \sum_{j=k+1}^{n} |t_{kj}|^2 .$$

This means that $\sum_{j=k+1}^{n} |t_{kj}|^2 = 0$ and thus

$$t_{kj} = 0, \quad j = k+1, \ldots, n .$$

As k ranges from 1 to n, we conclude that T is diagonal. □

The traces of two unitarily equivalent matrices are invariant under the unitary equivalence. The following theorem shows $tr(A^*A) = tr(B^*B)$.

Theorem 4. *Let $A \in M_n$ and $B \in M_n$ be unitarily equivalent, then*

$$\sum_{i,j=1}^{n} |a_{ij}|^2 = \sum_{i,j}^{n} |b_{ij}|^2 \, .$$

Proof: Let $A = U^*BU$ where U is unitary. Then

$$tr(A^*A) = tr(U^*B^*UU^*BU) = tr(U^*B^*BU) = tr(B^*B) \, .$$

As $\sum_{i,j=1}^{n} |a_{ij}|^2 = tr(A^*A)$, the assertion follows. □

2.2.5 Congruence Transformation

Definition 14 (Congruent matrices)**.** *Two matrices $A, B \in M_n$ of the same order are congruent if there exist a nonsingular matrix $S \in M_n$ such that*

$$A = SBS^*.$$

Every $n \times n$ Hermitian matrix is congruent to a unique matrix which has the simple form

$$\begin{pmatrix} \pi & & \\ & \nu & \\ & & \delta \end{pmatrix}$$

where π, ν, δ represent the positive, negative and zeroes eigenvalues of A respectively. A matrix with this kind of a simple form is called an inertia matrix.

Therefore, employing the congruence concept the set of all $n \times n$ Hermitian matrices can be partitioned into equivalence classes.

Definition 15. *The inertia of a Hermitian matrix $A \in M_n$ is the ordered triple*

$$In(A) = (i_+(A), i_-(A), i_0(A))$$

where $i_+(A)$ is the number of positive eigenvalues of A, $i_-(A)$ is the number of negative eigenvalues of A, and $i_0(A)$ is the number of zero eigenvalues of A.

The rank of A is the quantity $i_+(A) + i_-(A)$.

Congruence is an equivalence relationship which can be summarized as follows:

- A is congruent to itself *(reflexivity)*
- If A is congruent to B, then B is congruent to A *(symmetry)*
- If A is congruent to B and B is congruent to C, then A is congruent to C *(transitivity)*

Theorem 5. *Let $A, B \in M_n$ be Hermitian; A and B are congruent if and only if they have the same inertia.*

Proof: See [20] theorem 4.5.8 □

2.2.6 Properties of Hermitian and Symmetric Matrices

Our prime objective in this section is to list the properties of the crucial topic of the eigenvalues and the eigenvectors of Hermitian matrices. Hermitian matrices arise naturally in many engineering and physics problems.

The importance of Hermitian and symmetric matrices stems from their unique properties which significantly simplify our efforts to compute their eigenvalues and eigenvectors. For example, Schur decomposition 2.2.4 can be directly employed in order to diagonalize symmetric and Hermitian matrices. Furthermore, given any Hermitian or symmetric matrix, an orthonormal basis $\{x_1, x_2, \dots, x_n\}$ can be found such that $x_i, (i = 1, \dots, n)$ is an eigenvector of A and hence A is unitarily diagonalizable. We denote by H_n the set of all $n \times n$ Hermitian matrices.

Some of the important properties of Hermitian matrices are listed below, the proof of all is trivial.

Proposition 3. *1. Hermitian matrices are normal matrices.*

2. If $A \in M_n$ is Hermitian, then A^k is also Hermitian for all $k = \{1, 2, \dots, n\}$.

3. The sum of two Hermitian matrices is Hermitian.

4. The products of two Hermitian matrices is Hermitian only when they commute.

5. Any matrix $A \in M_n$ can be written as $A = H(A) + S(A)$ where $H(A) = \frac{1}{2}(A + A^*)$ is the Hermitian part of A, and $S(A) = \frac{1}{2}(A - A^*)$ is the skew-Hermitian part of A

Theorem 6. *If Hermitian matrix A has distinct eigenvalues, then its eigenvectors form a unitary matrix.*

Proof: From

$$Ax = \lambda x, \qquad Ay = \mu y$$

$\lambda \neq \mu$, we obtain

$$(y, Ax) = \lambda(y, x), \qquad (x, Ay) = \mu(x, y)\,.$$

Since $(x, Ay) = (Ax, y) = (y, Ax)$, we get $(\lambda - \mu)(x, y) = 0$ and so $(x, y) = 0$. □

The following theorem shows that the eigenvalues of Hermitian matrices are necessarily real. Whereas the eigenvectors of Hermitian matrices are generally complex. However, the converse is not true; a matrix with real eigenvalues is not necessarily Hermitian. This theorem has very important implications such as the ability to order eigenvalues in any prescribed order.

The fact that the eigenvalues of Hermitian matrices are always real, and that they can be ordered in any prescribed order, is extremely useful for the computation of extremal eigenvalues and the spectrum. Throughout this thesis we shall order the eigenvalues of Hermitian matrices in a non increasing order.

Theorem 7. *Let $A \in M_n$ be Hermitian, then all the eigenvalues of A are necessarily real.*

Proof: Let $Ax = \lambda x$ with x, λ being eigenvector and eigenvalue of A respectively. Then

$$x^*Ax = \lambda x^*x, \quad \text{i.e.,} \quad \lambda = \frac{x^*Ax}{x^*x}\,.$$

Now $\overline{(x^*Ax)} = (x^*Ax)^* = x^*A^*x = x^*Ax$, so x^*Ax is real and thus λ is real. □

One of the important properties of Hermitian matrices is that Hermitian matrices can be diagonalized using unitary similarity. The following theorem sheds light on this.

Theorem 8. *Let $A \in M_n$ be Hermitian, then A is unitarily similar to a diagonal matrix.*

Proof: According to Proposition 3 Hermitian matrices are normal; furthermore, a normal matrix is unitarily similar to a diagonal matrix. $\square$

2.2.7 Unitary Matrices

Unitary matrices are important in numerous areas such as quantum mechanics, quantum computing, the study of complex covariance matrices. Unitary matrices are used to transform matrices into canonical forms, see [12] p.70, [20] p.66-72. For further details see section 2 in the Schur's unitary triangularization. The elementary unitary matrices that are Hermitian such as Householder are used for the computation of the QR factorization which is the subject of the Section 3.1.3. Also they can be used for orthogonal transformation of a square matrix $A \in M_n$ into Hessenberg form. We have seen earlier that every complex matrix A is unitarily similar to an upper triangular matrix (or diagonal) with diagonal entries equal to the eigenvalues of A. Also, unitary matrices can be used to determine whether two matrices are unitarily equivalent.

Unitary matrices are complex square matrices which have columns or rows that form an orthonormal set of vectors. If we denote a^i the i^{th} column of $A,\ i = 1, \cdots, n$, then $A^*A = I$ will mean:

$$a^{i*}a^{(j)} = \begin{cases} 0 & \text{if } i \neq j, \\ \\ 1 & \text{if } i = j \end{cases}$$

Likewise the rows of A^* form an orthonormal set.

The determinant of a unitary matrix has an absolute value of 1:

$$|det(U)|^2 = det(U^*)det(U) = det(|I|) = 1$$

The product of two unitary matrices of the same order is unitary. Also the sum of two unitary matrices of the same order is also unitary.

The inverse of a unitary matrix is equal to its conjugate transpose.This is important because its inverse can be found easily by finding its conjugate transpose. A unitary matrix which all of its elements are real is referred to as an orthogonal matrix.

If a square matrix $A \in M_n$ is unitary then a unitary transformation $y = Ax$ preserves the values of inner product:

$$y = Ax \text{ then } y^*y = (Ax)^*Ax = x^*A^*Ax = x^*Ix = x^*x\,.$$

Furthermore, the magnitude of any $x \in \mathbb{C}^n$ is also preserved, see [20] p.66-72, [43] p.26.

Diagonal matrices are better suited in this purpose as there are large varieties of upper triangular matrices. Two matrices are similar if both of them are similar to the same diagonal matrix. Which means the diagonal elements of the two matrices are same in terms of multiplicities. However, there is still the difficulty that not every complex matrix is similar to a diagonal matrix.

2.2.8 Orthogonal Matrices

Orthogonal matrices are crucial for matrix theory and warrant a special consideration. In the theory of matrix analysis it is very useful to know whether a certain type of transformation guarantees the preservation of useful property. The different transformations preserve different properties. For example similarity transformation preserves eigenvalues. Likewise, congruence transformation preserves the nice properties of symmetry and Hermitian. The orthogonal matrices preserve the inner product of vectors such as the transformation of matrices that represent rotation of a plane about the origin and a reflection of a plane, see [12] p.70, [39] p.333, [20] pp.71-72.

The following properties of orthogonal matrices is similar to the above definition and can also be used as alternative definitions.

The fact that $A \in M_n$ is unitary is equivalent to each of the following:

- $\|Ax\| = \|x\|$ for all $x \in \mathbb{C}^n$
- the angle between two vectors x, y defined as $cos\Theta = \frac{\langle x,y\rangle}{\|x\|_2\|y\|_2}$ is equal to the angle between $Ax, Ay, \forall x, y \in \mathbb{C}^n$
- $\|AB\|_2 = \|B\|_2$ for all $A \in M_n$
- $\|BA\|_2 = \|B\|_2$ for all $A \in M_n$
- The condition number $\kappa(A)$ of an orthogonal matrix is one. To see this note that

$$\|A\| = 1 \text{ and } \kappa(A) = \|A\|\|A^{-1}\|; \text{ then } \|A\|\|A^{-1}\| = 1\,.$$

It is possible to convert $m \times n$ nonsingular matrix to a diagonal matrix by applying a sequence of unitary transformations. These orthogonalization processes include Householder reflection and Given's rotation. These orthogonalization processes are discussed in more detail in section 3.

2.3 Spectral Decomposition

2.3.1 Spectral Theorem of Hermitian and Symmetric Matrices

One of the most important properties of Hermitian matrices is the spectral theorem for it guarantees a set of complete orthonormal eigenvectors. The Spectral theorem for Hermitian matrices provides the possibility of diagonalizing Hermitian matrices by unitary similarity transormation. The diagonal form provides us with the eigenvalues and the computation of eigenvectors becomes relatively easy, see [20] p.104. All too often one might be concerned to compute the largest or the smallest eigenvalue of a matrix. The spectral theorem simplifies the process by making it possible to determine the spread of the eigenvalues of a Hermitian matrix. To study the spectral theorem for Hermitian matrices we utilize the Schur's unitary triangularization which is the subject for the following section.

Theorem 9. *Let $A \in M_n$ be Hermitian then there is a unitary matrix $U \in M_n$ and a real diagonal matrix $\Lambda \in M_n$ such that $A = U\Lambda U^*$ and the diagonal entries of Λ are the eigenvalues of A, i.e., $\Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$.*

Proof: Let $A \in M_n$, then $\exists$ unitary matrix U such that $U^*AU = T$ (upper triangular), i.e., $A = UTU^*$. If we take the complex conjugate transpose of A we get

$$A = A^* = (UTU^*)^* = U^{**}T^*U^* = UTU^*,$$

thus, $T^* = T$, so T is real diagonal. Let us now show that the elements of T are the eigenvalues of A. We know that $A = UTU^*$, hence $AU = UT$ and thus $Au_i = t_iu_i$, where u_i is the $i-th$ column of U and t_i the $i-th$ diagonal element of T. Since $u^*u = 1$, then t_i $(i = 1, 2, \ldots, n)$ must be the eigenvalues of A. □

2.3.2 Rayleigh-Ritz

The largest and the smallest eigenvalues of a Hermitian matrix can be found by using Rayleigh-Ritz iteration. This is an important optimization problem which finds the minimum and the maximum of the scalar quantity $\min\{x^*Hx : x^*x \neq 0, x \in \mathbb{C}^n\}$. Hence, Rayleigh-Ritz ratio is used to identify the maximum and the minimum eigenvalues of a Hermitian matrix.

Theorem 10 (Rayleigh-Ritz)**.** *Let $H \in M_n$ be Hermitian. Assuming that the eigenvalues of H are ordered as*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq \lambda_n$$

we have

$$\lambda_n x^*x \leq x^*Hx \leq \lambda_1 x^*x \quad \forall x \ \in \mathbb{C}^n,$$

$$\lambda_{max} = \lambda_1 = \max_{x \neq 0} \frac{x^*Hx}{x^*x} = \max\{x^*Hx : x^*x = 1\},$$

and

$$\lambda_{min} = \lambda_n = \min_{x \neq 0} \frac{x^*Hx}{x^*x} = \min\{x^*Hx : x^*x = 1\}.$$

Proof: To prove this, Theorem 9 is used. Since H is Hermitian, there exist a unitary matrix $U \in M_n$ such that $H = U\Lambda U^*$ where the eigenvalues of H are the diagonal entries of matrix Λ. For any vector $x \in \mathbb{C}^n$ with $||x|| = 1$ we get

$$x^*Hx = x^*U\Lambda U^*x = \sum_{i=1}^{n} \lambda_i |(U^*x)_i|^2\,,$$

therefore

$$\lambda_n \sum_{i=1}^{n} |(U^*x)_i|^2 \leq x^*Hx \leq \lambda_1 \sum_{i=1}^{n} |(U^*x)_i|^2$$

and, as U is unitary and $(U^*x)^*U^*x = x^*x$, we get the first claim.

Now assume that x is the eigenvector corresponding to λ_1, then $x^*Hx = \lambda_1 x^*x$, and analogously for the smallest eigenvalue λ_n. Hence the inequalities are sharp and we are done. □

Example 2. Here is a numerical illustration:

$$A = \begin{pmatrix} 11 & -3 & 5 & -8 \\ -3 & 11 & -5 & -8 \\ 5 & -5 & 19 & 0 \\ -8 & -8 & 0 & 16 \end{pmatrix}.$$

The matlab function $[V, e] = eig(A)$ computes respectively the eigenvalues and eigenvectors of a matrix A. Using this function we get the orthogonal matrix formed from the normalized eigenvectors of A and the eigenvalues of A written in descending order.

$$V = \begin{pmatrix} -0.5774 & 0.5774 & -0.4082 & -0.4082 \\ -0.5774 & -0.5774 & -0.4082 & 0.4082 \\ 0 & -0.5774 & 0 & -0.8165 \\ -0.5774 & 0 & 0.8165 & 0 \end{pmatrix}.$$

The eigenvalues of A written in non-increasing order are $[24, 24, 9, 0]$

$$V' * A * V = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -0.0000 & 9.0000 & 0.0000 & -0.0000 \\ 0 & 0.0000 & 24.0000 & 0.0000 \\ -0.0000 & -0.0000 & 0.0000 & 24.0000 \end{pmatrix}$$

The eigenvector of A corresponding to the above eigenvalues are

iterations	*eigenvectors*
v1	-0.5774, -0.5774, 0 ,-0.5774
v2	0.5774,-0.5774,-0.5774,0
v3	-0.4082,-0.4082 ,0 ,0.8165
v4	-0.4082,0.4082,-0.8165,0

The minimum is achieved when eigenvector is equal to v_1 which corresponds to l_1

iterations	l1	l2	l3	l4
eigenvalues	(v1'*A*v1) = 0	(v2'*A*v2) = 9.0	(v3'*A*v3) = 24.0	(v4'*A*v4) = 24.0

The maximum is achieved when eigenvector is equal to v_4 which corresponds to l_4 : $l_4 = v_4' * A * v_4 \ \Rightarrow l_4 = 24.0$.

2.3.3 Courant-Fischer Minimax

The Courant-Fischer minimax theorem is a very useful tool for analyzing the eigenvalues of Hermitian matrices.

Let $H \in M_n$ be a Hermitian matrix with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. We have seen in the Raylegh-Ritz Theorem 10 that the extreme eigenvalues can be written in the form of so-called Rayleigh quotient

$$\lambda_1 = \frac{x^*Hx}{x^*x}, \qquad \lambda_n = \frac{y^*Hy}{y^*y}$$

with appropriate x and y. The Courant-Fisher minimax theorem deals with the intermediate eigenvalues.

Theorem 11 (Courant-Fischer Minimax)**.** *Let $H \in M_n$ be Hermitian, and let $k \in \mathbb{R}$ with $1 \leq k \leq n$, then*

$$\lambda_k(H) = \max_{W:\, dim(W)=k} \ \min_{x\in W,\, x\neq 0} \frac{x^*Hx}{x^*x} \tag{2.1}$$

$$\lambda_k(H) = \min_{W:\, dim(W)=n-k+1} \ \max_{x\in W,\, x\neq 0} \frac{x^*Hx}{x^*x}\,. \tag{2.2}$$

Proof: We will only prove the second part, the first one would be analogous. Since H is Hermitian, it has an orthonormal basis of eigenvectors $x_1, \ldots, x_n$ corresponding to $\lambda_1, \ldots, \lambda_n$. Let $\hat{W}_k = \operatorname{span}(x_{n-k+1}, \ldots, x_n)$. Then, for any $x \in \hat{W}_k$ with $x = \sum_{i=n-k+1}^{n} \alpha_i x_i$ and $x^*x = \sum_{i=n-k+1}^{n} |\alpha_i|^2 = 1$,

$$x^*Hx = \sum_{i=n-k+1}^{n} \lambda_i |\alpha_i|^2 \leq \lambda_k \sum_{i=n-k+1}^{n} |\alpha_i|^2 = \lambda_k$$

so that the quantity on the right-hand side of (2.2) is no greater than λ_k. On the other hand, if $\hat{W} = \operatorname{span}(x_1, \ldots, x_{n-k+1})$, then for any $x = \sum_{i=1}^{n-k+1} \alpha_i x_i$ in $\hat{W}_k$, with $x^*x = 1$, we have

$$x^*Hx = \sum_{i=1}^{n-k+1} \lambda_i |\alpha_i|^2 \geq \lambda_k\,.$$

But $\dim \hat{W}_k = k$, so that the intersection of $\hat{W}_k$ with any $(n-k+1)$-dimensional subspace W_k must be at least one-dimensional. Hence, the right-hand side of (2.2) is at least as large as λ_k. □

Eigenvalues of the Sum of Two Hermitian Matrices

The maximum of the sum of two continues functions is less and equal to the sum of the individual maximum of each function. This idea which is presented below is used to illustrate the extremum of eigenvalues of the sum of two matrices.

Theorem 12. *Let $f, g : X \to R$ be continuous functions. Then*

$$\max_{x \in X}(f(x) + g(x)) \leq \max_{x \in X} f(x) + \max_{x \in X} g(x) \, .$$

Proof: Let $x^* \in R$ such that

$$f(x^*) + g(x^*) = \max_{x \in X} \ (f(x) + g(x)),$$

then

$$f(x^*) \leq \max_{x \in X} \ f(x), \qquad g(x^*) \leq \max_{x \in X} \ g(x),$$

so

$$\max_{x \in X} \ (f(x) + g(x)) = f(x^*) + g(x^*) \leq \max_{x \in X} \ f(x) + \max_{x \in X} \ g(x) \quad \Box$$

To investigate the relationship between the eigenvalues of two Hermitian matrices we use the result of the previous two theorems.

Theorem 13. *Let A and $B \in H_n$ then*

$$\lambda_1(\frac{A+B}{2}) \leq \frac{1}{2}[\lambda_1(A) + \lambda_1(B)]$$

$$\lambda_n(\frac{A+B}{2}) \geq \frac{1}{2}[\lambda_n(A) + \lambda_n(B)]\,.$$

Proof: From the Rayleigh-Ritz theorem, we have

$$\lambda_1(A) = \max[x^*Ax : \|x\|_2 = 1; x \in \mathbb{C}^n].$$

$$\lambda_1(B) = \max[x^*Bx : \|x\|_2 = 1; x \in \mathbb{C}^n].$$

This implies that (by Theorem 12)

$$\begin{aligned}
\lambda_1(\frac{A+B}{2}) &= \max[x^*\frac{(A+B)}{2}x : \|x\|_2 = 1; x \in \mathbb{C}^n] \\
&\leq \max[\frac{x^*Ax}{2} + \frac{x^*Bx}{2} : \|x\|_2 = 1; x \in \mathbb{C}^n] \\
&\leq \max[\frac{x^*Ax}{2} : \|x\|_2 = 1] + \max[\frac{x^*Bx}{2} : \|x\|_2 = 1] \\
&= \frac{1}{2}[\max(x^*Ax : x^*x = 1) + \max(x^*Bx : x^*x = 1)] \\
&= \frac{1}{2}[\lambda_1(A) + \lambda_1(B)]
\end{aligned}$$

The second part of the theorem is proven in a similar way. □

2.3.4 Weyl's Perturbation Theorem of Hermitian Matrices

This subsection treats the eigenvalues of the sum of two Hermitian matrices. Weyl's Theorem and the weaker version of Weyl's theorem will be used to give lower and upper bounds for the eigenvalues of $A + E$. Furthermore, Weyl's theorem sheds light on how to derive the bounds of the distance between the eigenvalues of a Hermitian matrix A and those of a skew-Hermitian matrix B.

If $A \in M_n$ is a Hermitian matrix and we add a positive definite matrix to it, then all the eigenvalues of A will increase. This is the subject of this theorem.

Theorem 14 (Weyl's theorem)**.** *Let $A, E \in H_n$ then for all $k : 1 \leq k \leq n$ it holds that*

$$\lambda_k(A) + \lambda_n(E) \leq \lambda_k(A+E) \leq \lambda_k(A) + \lambda_1(E).$$

Proof: For any $x \in \mathbb{C}^n$, $x \neq 0$, we get

$$\lambda_n(E) \leq \frac{x^*Ex}{x^*x} \leq \lambda_1(E).$$

For any $k = 1, 2, \ldots, n$ we get

$$\begin{aligned} \lambda_k(A+E) &= \min_{W:\, dim(W)=n-k+1} \; \max_{x \in W,\, x \neq 0} \frac{x^*(A+E)x}{x^*x} \\ &= \min_{W:\, dim(W)=n-k+1} \; \max_{x \in W,\, x \neq 0} \left(\frac{x^*Ax}{x^*x} + \frac{x^*Ex}{x^*x} \right) \\ &\geq \min_{W:\, dim(W)=n-k+1} \; \max_{x \in W,\, x \neq 0} \left(\frac{x^*Ax}{x^*x} + \lambda_n(E) \right) \\ &= \lambda_k(A) + \lambda_n(E)\,. \end{aligned}$$

The upper bound can be proven analogously. □

An immediate consequence of the Weyl's theorem is the following theorem.

Theorem 15. *Let $A, B \in M_n$ be Hermitian and let $B \geq 0$. Then $\lambda_k(A) \leq \lambda_k(A+B)$ for any $k : 1 \leq k \leq n$.*

2.3.5 Interlacing Eigenvalues Theorem

If matrix A is perturbed by a rank one matrix the result is that the eigenvalues of A are shifted in a regular way. The eigenvalues of $(A + xx^*)$ interlace with those of A. There is one eigenvalue of A between each pair of odd or even eigenvalues of $(A + xx^*)$, see [20] p.182, [12] p.411.

Theorem 16. *Let $A \in H_n$ and $x \in \mathbb{C}^n$, then*

$$\lambda_{i+1}(A + xx^*) \leq \lambda_i(A) \leq \lambda_i(A + xx^*)\,,$$

where $i = 1, 2, \ldots, n-1$.

Proof: There exists a subspace X of dimension $i+1$ such that

$$\lambda_{i+1}(A + xx^*) = \min_{y \in X}\{\frac{y^*(A + xx^*)y}{y^*y} : y \neq 0\}.$$

Let $\hat{X} = X \bigcap \{y : y^*x = 0\}$; evidently, the dimension of $\hat{X}$ is either $i+1$ or i. Let $dim(\hat{X}) = j$. Now replace X with $\hat{X}$ to get

$$\begin{array}{rcl} \lambda_{i+1}(A + xx^*) & = & \min\limits_{y \in X}\{\frac{y^*(A + xx^*)y}{y^*y} : y \neq 0\} \\ & \leq & \min\limits_{\hat{X}} y \in \hat{X}\{\frac{y^*(Ay + xx^*y)}{y^*y} : y \neq 0\}. \end{array}$$

Since $\hat{X} \subseteq X$ and $y^*x = 0$ we get

$$\begin{array}{rcl} \lambda_{i+1}(A + xx^*) & = & \min\limits_{y \in \hat{X}}\{\frac{y^*Ay}{y^*y} : y \neq 0\} \\ & \leq & \max\limits_{Z \subset \mathbb{R}^n,\ dim(Z)=j}\ \min\limits_{y \in Z}\{\frac{y^*Ay}{y^*y} : y \neq 0\} \\ & = & \lambda_j(A). \end{array}$$

Now since $j = i$ or $i+1$, $\lambda_j(A) \leq \lambda_i(A)$. The other bound is proved analogously.

□

Example 3.

$$A = \begin{pmatrix} 9 & 2 & 1 \\ 2 & 6 & 3 \\ 1 & 3 & 11 \end{pmatrix};\ x = \begin{pmatrix} 0.4243 \\ 0.5657 \\ 0.7071 \end{pmatrix}$$

Let $C = A + xx^T$

$$C = \begin{pmatrix} 9.1800 & 2.2400 & 1.3000 \\ 2.2400 & 6.3200 & 3.4000 \\ 1.3000 & 3.4000 & 11.5000 \end{pmatrix}$$

$$\alpha = eig(A);\ \gamma = eig(C)$$

$$\alpha = \{\ 13.1823 \quad 8.6539 \quad 4.1638\ \}$$

$$\gamma = \{\ 14.1620 \quad 8.6584 \quad 4.1796\ \}$$

Notice that $\alpha_3 \leq \gamma_3 \leq \alpha_2 \leq \gamma_2 \leq \alpha_1 \leq \gamma_1$

2.3.6 Gerschgorin's Theorem

The main function of Gerschgorin's theorem lies in locating the eigenvalues of a matrix. Similarly it can be applied on locating the field of values of a matrix (which is the subject of Chapter 4). Moreover, Gerschgorin's theorem has many numerical applications, for example it can be used for the testing of whether a matrix is invertible.

Definition 16. *$A \in M_n$ is diagonally dominant if*

$$a_{ii} > \sum_{j \neq i} |a_{ij}|, i = 1, \ldots, n.$$

Proposition 4. *If $A \in M_n$ is diagonally dominant then it is invertable.*

Proof: A is invertible when $Ax \neq 0$, $\forall x \neq 0$. Take any $x \in \mathcal{C}^n, x \neq 0$, and choose a component i such that $|x_i| = \max_j |x_j|$. Then

$$\begin{aligned}
|(Ax)_i| &= |\sum_{j=1}^{n} a_{ij}x_j| \\
&= |a_{ii}x_i + \sum_{j\neq i} a_{ij}x_j| \\
&\geq |a_{ii}x_i| - |\sum_{j\neq i} a_{ij}x_j| \\
&\geq |a_{ij}||x_i| - \sum_{j\neq i} |a_{ij}||x_i| \\
&= |x_i|(|a_{ii}| - \sum_{j\neq i} |a_{ij}|)\,.
\end{aligned}$$

From the diagonal dominance of A, $(|a_{ii}| - \sum_{j\neq i} |a_{ij}|)$ is positive; furthermore, $|x_i| > 0$, so the last term is strictly positive and thus $Ax \neq 0$. □

Theorem 17 (Gerschgorin)**.** *Let $A \in M_n$ and let*

$$D_i = \{z : |z - a_{ii}| \leq \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|\}, \quad i = 1, 2, \ldots, n\,.$$

Then

- *all n eigenvalues of A lie in the union of discs $\bigcup_{i=1}^{n} D_i$;*
- *if k of the discs are disjoint from the other $n-k$ discs, then the k discs contain exactly k eigenvalues of A.*

Proof:

- Assume by contradiction that there exist an eigenvalue λ such that $\lambda \notin$

$\bigcup_{i=1}^{n} D_i$. Then

$$|a_{ii} - \lambda| > \sum_{\substack{i \neq j \\ j=1}}^{n} |a_{ij}|; \ i = \{1, \dots, n\}|$$

This implies that $(A - \lambda I)$ is diagonally dominant, which also means $(A - \lambda I)$ is invertible. Hence λ is not an eigenvalue of A.

- Using the fact that eigenvalues are continuous family of functions, let

$$A(t) = \underbrace{D}_{\text{diagonal part}} + \underbrace{tN}_{\text{offdiagonal part}}$$

for $t \in [0, 1]$. So $A(0) = D$ corresponding to eigenvalues a_{ii} of $A(t)$. As t increases from 0 to 1, the radius of the discs increase up to $A(1) = A$. The k discs are still disjoint from the other $n - k$ eigenvalues, since the k eigenvalues in the union of k discs can not separate from the union by the continuity argument. □

Gerischgorin theorem may be applied to testing for the important property of positive definiteness of a matrix. The next corollary follows immediately and describes this application.

Corollary 1. *If $A = [a_{ij}] \in M_n$ is Hermitian and strictly diagonally dominant, and if $a_{ii} > 0, \ \forall (i = 1, 2, \cdots, n)$, then A is positive definite.*

3. MATRIX DECOMPOSITION

Matrix decomposition is the operation of factorizing matrices into some canonical forms. This entails to represent a matrix as the product of 2 or more matrices with special properties which exposes the nice properties of the matrix as well as simplifying the solution of linear algebra problems.

The matrices with special forms such as triangular and diagonal matrices play and important role in computing the eigenvalues and eigenvectors. The first part tackles the issue of decomposition related to solving systems of linear equations. The second part of matrix decomposition treats the decomposition related to eigenvalues. Some of the theorems that underpin the uses of these special forms are discussed in this section. The notion of similarity and congruence transformations are revisited and applied to matrix decomposition.

3.1 Decomposition Related to Solving Systems of Linear Equations

Triangular forms of matrices are quite important for the solutions of general linear systems $Ax = b$. Triangular or diagonal matrices simplify significantly the solution of systems of equations. The unknown variables can be found one at a time by either forward or backward substitution, see [12] p. 92, [20] p. 158-165.

A matrix $A = [a_{ij}] \in M_n$ is said to be upper triangular if $a_{ij} = 0$ whenever $j < i$, in other words all the elements below the diagonal are zero. Similarly A is said to be lower triangular if $a_{ij} = 0$ whenever $i < j$, which means all the elements above the diagonal are zero. A triangular matrix is singular if and only if one or more elements in the diagonal is zero. A square matrix A can be written as the product of simpler matrices such as product of lower and upper triangular matrices $A = LU$ with the same dimension as A. The following section is dedicated to this type of matrix decomposition.

3.1.1 Block LU Factorization

In most cases if $A \in M_n$ is nonsingular, then A can be factored as LU with $L \in M_n$ block lower triangular and can be constructed in a such a way that it has identity matrices I_k on the diagonal. Matrix $U \in M_n$ is block upper triangular with non-zero (block) diagonal entries.

However, not every matrix has LU factorization as obvious from the following example from Horn and Johnson [20].

Example 4.

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Since $a_{11} = 0$ this matrix cannot be factored as LU. In block matrix terms this corresponds to the leading principal sub-matrix $A_{11} = 0$, as a result its determinant $det(A)$ will be zero.

Suppose $A \in M_n$ can be block LU factorized as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \; L = \begin{pmatrix} I_1 & 0 \\ L_{21} & I_2 \end{pmatrix}, \; U = \begin{pmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{pmatrix},$$

where I_1 and I_2 are identity matrices of appropriate dimensions. To solve a system of equations $LUx = b$ using the newly factorized matrix LU:

1. first substitute y with Ux and solve the system $Ly = b$ for y using forward substitution.

2. secondly solve $Ux = y$ for x using backward substitution.

In the case when A is positive definite the LU factorization can have the special form $A = G^*G$. This special factorization is called the Cholesky factorization and is the subject for discussion in section 3.1.2.

Schur complement is related to the block LU factorization. Consider the 2-block factorization from the above example. As $A = LU$, we get that

$$\begin{aligned} A_{11} &= I_1U_{11} = U_{11} \\ A_{12} &= I_1U_{12} = U_{12} \\ A_{21} &= L_{21}U_{11}, \quad \text{i.e. } L_{21} = A_{21}U_{11}^{-1} = A_{21}A_{11}^{-1} \\ A_{22} &= L_{21}U_{12} + U_{22} = L_{21}A_{12} + U_{22}\,. \end{aligned}$$

After substitution, we get

$$A_{22} = A_{21}A_{11}^{-1}A_{12} + U_{22}$$

and thus

$$U_{22} = A_{22} - A_{21}A_{11}^{-1}A_{12},$$

which is the Schur Complement of A_{11} in A. Thus we have

$$A = LU = \begin{pmatrix} I & 0 \\ A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix}$$

Taking determinants we have

$$det(A) = det(L)det(U) = det(A_{11})det(A_{22} - A_{21}A_{11}^{-1}A_{12})\,.$$

We conclude this section with an algorithm for the general LU-factorization. Here we use pseudo code in order to simplify the reading.

Algorithm 1 (Write $A = LU$ in terms of its entries)**.** *Take the diagonal elements of* L *to be* $1's$

```
for $k = 1, \dots, n$ do
    $u_{kk} = a_{kk} - \Sigma_{m=1}^{k-1} l_{km}u_{mk}$
    for $j = k, \dots n$ do
        $u_{kj} = a_{kj} - \Sigma_{m=1}^{k-1} l_{km}u_{mj}$
    end
    for $i = k+1, \dots n$ do
        $l_{ik} = (a_{ik} - \Sigma_{m=1}^{k-1} l_{im}u_{mk})/u_{kk})$
    end
end
```

3.1.2 Cholesky Factorization

A symmetric positive definite matrix $A \in M_n$ can be factorized into a lower triangular matrix $G \in M_{n\times n}$ and its transpose. The lower triangular matrix G is called Cholesky factor of the original positive definite matrix A and the diagonal entries are all positive. This factorization can be viewed as LU factorization where U is the transpose of L.

Theorem 18. *Let $A \in M_n$ be Hermitian and positive definite, then exists a unique lower triangular matrix $G \in M_{n\times n}$ with positive diagonal entries such that $A = GG^*$*

Proof: To find G, the elements of $G = [g_{ij}]$ are computed column by column starting from the top. Each element of $A = [a_{ij}]$ is written as a function of the inner product of i-th row of G and j-th column of G^*

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} g_{11} & 0 & \cdots & 0 \\ g_{21} & g_{22} & 0 & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ g_{n1} & \cdots & \cdots & g_{nn} \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ 0 & g_{22} & \cdots & g_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & g_{nn} \end{pmatrix}$$

The following two equations are used to find each element of G_{ij}.

To calculate the diagonal elements g_{ii} taking row by row.

$$a_{ii} = \sum_{ij}^{i} g_{ik}\bar{g_{ik}} \Rightarrow g_{ii} = \sqrt{a_{ii} - \sum_{j=1}^{i-1}(g_{ik})^2}\,.$$

To calculate the off diagonal elements g_{ij} taking row by row.

$$a_{ij} = \sum_{ij}^{i} g_{ik}g_{jk} \Rightarrow g_{ij} = (a_{ij} - \sum_{j=1}^{i-1} \frac{g_{ik}g_{jk}}{a_{ii}}) . \quad \square$$

Cholesky factorization is very useful and has many applications such as solving systems of linear equations see [20] p.407, [43] p.229 [12] p.141. Suppose $A = GG^T$ where G is lower triangular and $g_{ii} \neq 0$. Then we can solve the system of linear equations $Ax = b$ where matrix A is positive definite by writing the system $GG^T x = b$. First let $G^T x = y$ and solve $Gy = b$ for y by forward substitution. This operation requires only $O(n^2)$ flops. Once found the values of vector y it is easy to compute vector x from $G^T x = y$ by back substitution. This computation requires only $O(n^2)$ flops.

3.1.3 QR Factorization

We now consider QR factorization of a matrix $A \in M_{m,n},\ m \geq n$. This procedure generates

$$A = QR$$

where $Q \in M_{m,m}$ is orthogonal and $R \in M_{m,n}$ is upper triangular. To determine matrix Q, matrix A is factorized as product of elementary unitary Hermitian matrices using Householder reflections, or plane rotation using Given's transforms. For more detailed description see [12] p.351-355, [20] pp. 112-117.

QR Factorization comes in various forms depending on the type of matrix in hand.

Theorem 19. *Let $A \in M_{n,m}$ have full rank then there exist an orthogonal matrix $Q \in M_{m,m}$ and a nonsingular, upper-triangular matrix $R \in M_{m,n}$ such* *that $A = QR$. If $m = n$, Q is unitary.*

Proof: See [20, Thm.2.6.1]. □

Below are two of the most popular methods used for QR factorization

3.1.4 Computing the QR Factorization Using Householder Transformation

In this section a detailed description of QR factorization and its algorithm are provided. Several explanatory numerical examples are also included.

In the QR algorithm we repeatedly multiply A by an orthogonal matrix H_k. At the k-th step we require that all elements of $[H_k H_{k-1} \cdots Q_1 A$ below the diagonal become zero and the existing zeroes are preserved. This is achieved by means of the Householder transformation. Let $v \in R^n$, $\|v\|_2 = 1$.

Suppose that we are given a vector $x \in R^n$, $x \neq 0$ and want to find a matrix H such that Hx be a multiple of $e^{(1)} = (1, 0, \ldots, 0)^T$. Set $v = \dfrac{x - \|x\|_2 e^{(1)}}{\|x - \|x\|_2 \|e^{(1)}\|_2}$. The vector v is called a Householder vector and the matrix

$$H = I - 2vv^T$$

is called a Householder reflection and has the required property, namely,

$$Hx = \|x\|_2 e^{(1)}\,.$$

Algorithm 2 (Householder QR)**.** *Let $A \in M_{m,n}$. Set $R_0 = A$.*

```
for $j = 1, \ldots, n$ do
    compute the Householder vector $v_j$ for $(R_{j-1})_{j:m,j}$
    find the Householder matrix $H_j = I - 2v_j v_j^T$
    set $R_j = H_j^T R_{j-1}$
```

end

At the end, we get the desired matrices Q and R as

$$Q = H_n^T H_{n-1}^T \cdots H_1^T \quad \text{and} \quad R = R_n \,.$$

Example 5. *The following example factorizes the matrix*

$$A = \begin{pmatrix} 126 & 82 & -176 \\ 84 & 120 & 102 \\ 0 & -56 & 112 \\ 252 & 164 & -142 \end{pmatrix}$$

Iteration 1

$$x_1 = \begin{pmatrix} 126 \\ 84 \\ 0 \\ 252 \end{pmatrix} \quad \|x_1\|_2 = 294$$

$$y_1 = \|x_1\|_2 \; e^{(1)} = \begin{pmatrix} 294 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$v_1 = \frac{x_1 - y_1}{\|x_1 - y_1\|_2}; \quad v_1 = \begin{pmatrix} -0.5345 \\ 0.2673 \\ 0 \\ 0.8018 \end{pmatrix}$$

$$H_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - 2v_1 v_1^T$$

$$H_1 = \begin{pmatrix} 0.4286 & 0.2857 & 0 & 0.8571 \\ 0.2857 & 0.8571 & 0 & -0.4286 \\ 0 & 0 & 1.0000 & 0 \\ 0.8571 & -0.4286 & 0 & -0.2857 \end{pmatrix}$$

$$R_1 = H_1^T A = \begin{pmatrix} 294.0 & 210.0 & -168.0 \\ 0 & 56.0 & 98.0 \\ 0 & -56.0 & 112.0 \\ 0 & -28.0 & -154.0 \end{pmatrix}$$

Iteration 2

$$x_2 = \begin{pmatrix} 56.0 \\ -56.0 \\ -28.0 \end{pmatrix}; \quad \|x_2\|_2 = 84$$

$$y_2 = \begin{pmatrix} \|x_2\|_2 \\ 0 \\ 0 \end{pmatrix}$$

$$v_2 = \frac{x_2 - y_2}{\|x_2 - y_2\|_2} = \begin{pmatrix} -0.4082 \\ -0.8165 \\ -0.4082 \end{pmatrix}$$

$$H_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - 2v_2 v_2^T$$

$$H_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.6667 & -0.6667 & -0.3333 \\ 0 & -0.6667 & -0.3333 & -0.6667 \\ 0 & -0.3333 & -0.6667 & 0.6667 \end{pmatrix}$$

$$R_2 = H_2^T H_1^T A = \begin{pmatrix} -294.0 & -210.0 & 168.0 \\ 0 & -84.0 & -42.0 \\ 0 & 0 & -210.0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$Q = H_2 H_1$$

For comparison, below is the result of the testing of the above example using the MATLAB built in function $[Q, R] = qr(A)$

```
>> A=[126 82 -176;84 120 102;0 -56 112;252 164 -142]

A =

   126    82  -176
```

```
    84   120   102

     0   -56   112

   252   164  -142


>> [Q,R]=qr(A)

Q =

   -0.4286    0.0952    0.4762   -0.7619

   -0.2857   -0.7143   -0.5714   -0.2857

         0    0.6667   -0.6667   -0.3333

   -0.8571    0.1905   -0.0476    0.4762


R =

 -294.0000 -210.0000  168.0000

         0  -84.0000  -42.0000

         0         0 -210.0000

         0         0         0
```

3.1.5 Computing the QR Factorization Using Givens Rotation

Givens rotation is another method used for orthogonal matrix decomposition. The difference between Givens method and Householder method is that Householder is used when a group of elements in the same column is annihilated, whereas Givens method selectively eliminates the elements of a matrix one element at a time.

Givens rotation is a square rank-2 matrix of the following type

$$G(i,k,\theta) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

where $c = cos(\theta)$ and $s = sin(\theta)$ for some θ. The diagonal numbers c are on the ii and kk positions. Premultiplying by $G(i,k,\theta)$ amounts to counter-clockwise rotation of θ radians in the (i,k) coordinate plane. Given $x \in R^n$, its Givens rotation $y = G(i,k,\theta)^T x$ is

$$y_j = \begin{cases} cx_i - sx_k & j = i \\ sx_i + cx_k & j = k \\ x_j & j \neq i,k \end{cases} .$$

Hence, if we want to transform x such that, after the rotation, its k-th is equal to zero, we put

$$c = \frac{x_i}{\sqrt{x_i^2 + x_k^2}} \quad \text{and} \quad s = \frac{-x_k}{\sqrt{x_i^2 + x_k^2}} .$$

This leads to the following algorithm in which Givens rotations are used to annihilate, from below, elements under the k-th diagonal element for $k = 1, \ldots, n$.

Algorithm 3 (Givens QR)**.** *Let $A \in M_{m,n}$.*

```
for $j = 1, \ldots, n$ do
```

```
    for $j = m, m-1, \ldots, j+2, j+1$ do
        compute $c, s$ from $A_{i-1,j}$ and $A_{i,j}$
        set $A_{i-1:i,j:n} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}^T A_{i-1:i,j:n}$
    end
end
```

The following example uses the above algorithm to factorize the matrix A shown below. We will use the Matlab notation.

Example 6. .

```
A =  5      2      2

     3      6      3

     6      6      9

Q =

     1      0      0

     0      1      0

     0      0      1

Iteration 1:

r_1=sqrt(9+36)

r_1 = 6.7082    c=3/r_1 = 0.4472    s =6/r_1 = 0.8944

G_1 = [1 0 0;0 c -s;0 s c]

G_1 = 1.0000   0          0

          0    0.4472   -0.8944

          0    0.8944    0.4472

A_1=G_1'*A

A_1 =5.0000    2.0000    2.0000
```

```
    6.7082     8.0498    9.3915

         0    -2.6833    1.3416

Q_1 = Q*G_1'

Q_1 =    1.0000   0         0

          0       0.4472    0.8944

          0      -0.8944    0.4472

Iteration 2

r_2=sqrt(25+6.7082^2) = 8.3666

c=5/r_2 = 0.5976    s =6.7082/r_2 = 0.8018


G_2=[c -s 0;s c 0;0 0 1]

G_2 = 0.5976   -0.8018    0

     0.8018    0.5976     0

         0         0    1.0000

A_2 = G_2'*A_1

A_2 = 8.3666    7.6495    8.7252

      0.0000    3.2071    4.0089

         0     -2.6833    1.3416

Q_2 = Q_1*G_2'

Q_2 = 0.5976   0.8018   0

     -0.3586   0.2673   0.8944

     0.7171   -0.5345   0.4472

Iteration 3

r_3 = sqrt(3.2071^2+(-2.6833)^2) = 4.1816

c=3.2071/r_3 = 0.7670       s=-2.6833/r_3 = -0.6417
```

```
G_3=[1 0 0;0 c -s;0 s c]

G_3 =  1.0000    0      0

         0    0.7670    0.6417

         0   -0.6417    0.7670

A_3 = G_3'*A_2

A_3 =   8.3666    7.6495    8.7252

        0.0000    4.1816    2.2138

        0.0000    0.0000    3.6015

Q_3 = Q_2*G_3'

Q_3 =   0.5976    0.6149   -0.5145

       -0.3586    0.7789    0.5145

        0.7171   -0.1230    0.6860
```

The Matlab function `qr` produces the following result when applied to the same matrix.

```
 [Q,R]=qr(A)

Q =

   -0.5976    0.6149   -0.5145

   -0.3586   -0.7789   -0.5145

   -0.7171   -0.1230    0.6860

R =

   -8.3666   -7.6495   -8.7252

         0   -4.1816   -2.2138

         0         0    3.6015
```

There are differences in the sign of the elements of computed QR and the result from the Matlab built in function `qr`. This is due to the fact that the Matlab built in function `qr` starts the iteration from the bottom right hand corner and moves upward towards the top left hand corner A_{11}.

3.2 Factorization of Rectangular Matrices

This part is concerned with the important theorems of singular value decomposition and the polar decomposition, which extend the concept of matrix factorization to general, not necessarily square, complex matrices. The primary aim of this part is to introduce the singular value decomposition SVD and polar decomposition (PD). Both decompositions heavily depend on the concept of positive semidefiniteness. We conclude with the fundamental theorem of Hadamard's inequality.

3.2.1 Singular Value Decomposition (SVD)

Singular value decomposition has many applications such as statistical modeling, On many occasions one might want to work with rectangular matrices rather than square matrices. The singular value decomposition is well suited to this kind of situation since it can be applied to rectangular matrices as well as square matrices [20] pp.411-426.

A matrix is said to be singular if for a linear transformation a non zero input produces zero output. For singular value decomposition the input matrix V is multiplied by Σ to gain the corresponding output U. The nonzero outputs produced correspond to the rank of A. Therefore, the SVD is frequently used for determining the rank of a matrix and the relationship between rank of matrix A and that of neighbouring matrices.

Theorem 20. *Let $A \in M_{m,n}$ with $m \geq n$. Then there exist unitary matrices*

$U \in M_m$, $V \in M_n$ and a matrix $\Sigma \in M_{m,n}$ such that

$$A = U\Sigma V^*.$$

The matrix Σ with elements σ_{ij} has $\sigma_{ij} = 0$ for all $i \neq j$, and $\sigma_{11} \geq \sigma_{22} \geq \cdots \geq \sigma_{kk} > \sigma_{k+1,k+1} = \cdots = \sigma_{nn} = 0$. The numbers $\sigma_i := \sigma_{ii}$, $i = 1, \ldots, n$ are the nonnegative square roots of eigenvalues of AA^*. Moreover, the columns of the unitary matrix U are the eigenvectors of AA^*. Also, the columns of V form the eigenvectors of A^*A.

For a proof see [12, p.71] □

This makes A unitarily similar to Σ. Numbers σ_i are the *singular values* of A. The rank of A corresponds to the number k of non zero singular values of A. The U is the left singular vector whereas V^* is the right singular vector.

The singular values of a normal matrix A are the absolute values of the eigenvalues and the columns of V are the eigenvectors os A. If A is Hermitian, then all eigenvalues are real and the singular values are again their absolute values. If A is Hermitian and positive semidefinite, then the singular values are just the eigenvalues of A.

Usually it is not a good idea to compute the singular values of A by forming A^*A since this will inflate the condition number of the matrix. Instead, we can use the following theorem giving an equivalent characterization of the singular values.

Theorem 21. *Let $A \in M_{m,n}$ with $m \geq n$. Then σ_i $(i = 1, \ldots, n)$ are the singular values of A if and only if $\pm\sigma_i$, (i=1,...,n) and $m - n$ zeros are the* *eigenvalues of the matrix*

$$\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}.$$

For proof, see [20, Thm.7.3.7]

3.2.2 Polar Decomposition

Polar decomposition is an important and well studied matrix factorization. The polar decomposition states that for every $A \in M_n$ can be written as $A = PU$, where P is positive semidefinite with a rank equal to that of matrix A, whereas $U \in M_n$ is a unitary matrix.

Theorem 22. *Let $A \in M_{m,n}$ with $m \leq n$, then A may be written as $A = PU$ where $P \in M_n$ is positive definite and $U \in M_{m,n}$ has orthonormal rows.*

Proof: See [34].

Another variation of the above general case is when A is a square matrix, i.e., $m = n$. The following theorem considers this case. Its proof follows directly from the above theorem.

Theorem 23. *Let $A \in M_n$ then exist U unitary and P positive semidefinite such that $A = PU$. The matrix P is uniquely determined.*

We will finish this section with an important inequality for positive definite matrices.

Theorem 24 (Hadamard's Inequality)**.** *If $A \in M_n$ is positive definite then*

$$det(A) \leq \prod_{i=1}^{n} a_{ii}\,.$$

Proof: As A is positive definite, we have $a_{ii} \neq 0$ $(i = 1, \ldots, n)$. Define a diagonal matrix $D = \mathrm{diag}(a_{11}^{-1/2}, \ldots, a_{nn}^{-1/2})$. Clearly, $\det DAD \leq 1$ if and only if $\det A \leq a_{11}a_{22} \cdots a_{nn}$. We can thus assume that $a_{ii} = 1$ $(i = 1, \ldots, n)$. Hence we get

$$\det A = \prod_{i=1}^{n} \lambda_i \leq \left(\frac{1}{n} \sum_{i=1}^{n} \lambda_i \right)^n = \left(\frac{1}{n} tr\, A \right)^n = 1$$

where the inequality follows from the arithmetic-geometric mean inequality for nonnegative real numbers. □

3.3 Algorithms for Extreme Eigenvalues

This section addresses two methods used to determine the dominant eigenvalues and the associated eigenvectors of a matrix. The two are the power method and the inverse power method. Since these methods do not compute matrix decomposition they are suitable for large sparse matrices, which algebraic methods of computing eigenvalues and eigenvectors are not realizable. The main purpose of these methods is to find a single eigenvalue and its associated eigenvector. The methods used for computing all eigenvalues of a matrix include QR factorization, which is the subject of Section 3.1.3.

Before one starts finding the maximum or the minimum eigenvalue of a matrix it is prudent to first estimate the location of the largest eigenvalue by bounding the eigenvalues of the matrix. In other words, one needs to estimate the minimum and maximum values that the eigenvalues can acquire. Norms are quite suitable for this purpose. Let $A \in M_n$ and λ its eigenvalue such that $\rho(A) = |\lambda|$. Let further X be a matrix all columns of which are equal to the eigenvector associated with λ. If $|||\cdot|||$ is any matrix norm, then

$$|\lambda|\,|||X||| = |||\lambda X||| = |||AX||| \leq |||A|||\,|||X|||$$

and thus

$$|\lambda| = \rho(A) \leq |||A|||\,.$$

In words, the spectral radius of A is bounded above by any matrix norm of A. Particularly useful, as they are easy to compute, are the 1-norm $\|\cdot\|_1$, the infinity norm $\|\cdot\|_\infty$ and the Frobenius norm $\|\cdot\|_F$:

$$\begin{aligned}
\|A\|_1 &:= \max_{1\leq j\leq n}\sum_{i=1}^{m}|a_{ij}| \\
\|A\|_\infty &:= \max_{1\leq i\leq m}\sum_{j=1}^{n}|a_{ij}| \\
\|A\|_F &:= \sqrt{tr(A^*A)}\,.
\end{aligned}$$

To estimate the location of the largest eigenvalue of A, the above inequalities can be used. Other methods of estimating eigenvalues and their location include the trace of the matrix, which only tells us the sum of the eigenvalues, however, falls short of telling us the magnitude of the eigenvalues. Gerschgorin's discs allow us to find the n discs where the eigenvalues are found, however they do not give us the magnitude of individual eigenvalues.

Example 7. *In the following example we estimate the dominant eigenvalue* *using the above inequality:*

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{pmatrix}$$

The row sums of A are $4, 10, 20, 35$, respectively, hence, according to the above discussion, the spectral radius is bounded by 35 (the column sums are identical because of the symmetry). The Frobenius norm of this matrix is 26.4008. The actual largest eigenvalue is 26.3047 which is within the limits.

3.3.1 The Power Method

The power method computes the dominant eigenvalue of a square matrix and its corresponding eigenvector. The procedure is particularly useful when dealing with large sparse matrices since the computation does not require to store matrix A explicitly. The power method does not have many applications because its convergence is slow. We consider the following assumptions

- There is a single eigenvalue with greatest absolute value, which corresponds to the spectral radius of A.
- There is a set of linearly independent eigenvectors corresponding to the eigenvalues of A.

Theorem 25. *Suppose that $A \in M_n$ satisfies the two above assumptions. Then* *the sequence*

$$x^{(k+1)} = \frac{1}{(x^{(k)*}x^{(k)})^{1/2}} Ax^{(k)}, \quad k = 0, 1, 2, \ldots$$

converges to an eigenvector of A and the sequence $\lambda^{(k)} = x^{(k)*}Ax^{(k)}$ converges to the largest eigenvalue of A.

Proof: Let $x_1, \ldots, x_n$ be the (linearly independent) eigenvectors of A. Then any vector, say $x^{(0)}$, can be written as

$$x^{(0)} = \sum_{i=1}^{n} a_i x_i$$

with some a_i $(i = 1, \ldots, n)$. Multiplying the above equation by A^k gives

$$A^k x^{(0)} = A^k \sum_{i=1}^{n} a_i x_i = \sum_{i=1}^{n} a_i A^k x_i = \sum_{i=1}^{n} a_i \lambda_i^k x_i = a_n \lambda_n^k \left(x_n + \sum_{i=1}^{n-1} \frac{a_i}{a_n} \left(\frac{\lambda_i}{\lambda_n} \right)^k x_i \right).$$

We assumed that λ_n is the distinct largest (in absolute value) eigenvalue, hence the $\left(\frac{\lambda_i}{\lambda_n}\right)^k \to 0$ and thus, if $a_n \neq 0$ (and this follows from the first assumption), $A^k x^{(0)} \to a_n \lambda_n^k x_n$. □

The Euclidean norm can be, in fact, replaced by any other norm, the proof would not change.

Algorithm 4 (The Power Method Algorithm). *Let $A \in M_n$ satisfies the two above assumptions. Take any arbitrary vector $x^{(0)}$ with its largest component in absolute value equal to 1 and $\lambda^{(0)} = 0$ as approximation of an eigenvalue.*

```
For $k = 1, 2, \ldots$
    $x^{(k+1)} = \frac{Ax^{(k)}}{\|x^{(k)}\|}$
    $\lambda^{(k)} = x^{(k)*}Ax^{(k)}$
```

end

Example 8.

$$A = \begin{pmatrix} -261 & 209 & -49 \\ -530 & 422 & -98 \\ -800 & 631 & -144 \end{pmatrix}, \quad x^{(0)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Iteration 1

$$z^{(1)} = Ax^{(0)} = \begin{pmatrix} -101 \\ 206 \\ 313 \end{pmatrix}; \; x^{(1)} = \frac{z^{(1)}}{||z^{(1)}||_\infty} = \begin{pmatrix} -0.2603 \\ -0.5308 \\ -0.8065 \end{pmatrix}; \; \lambda^{(1)} = (x^{(1)})^T A x^{(1)} = 13.1907$$

Iteration 2

$$z^{(2)} = Ax^{(1)} = \begin{pmatrix} -3.4941 \\ -7.0295 \\ -10.6009 \end{pmatrix}; \; x^{(2)} = \begin{pmatrix} -0.2649 \\ 0.5329 \\ -0.8036 \end{pmatrix}; \; \lambda^{(2)} = 10.7462$$

Iteration 3

$$z^{(3)} = Ax^{(2)} = \begin{pmatrix} -2.8618 \\ -5.7361 \\ -8.6249 \end{pmatrix}; \; x^{(3)} = \begin{pmatrix} -0.2663 \\ 0.5338 \\ 0.8026 \end{pmatrix}; \; \lambda^{(3)} = 10.2159$$

Iteration 4

$$z^{(4)} = Ax^{(3)} = \begin{pmatrix} -2.7263 \\ -5.4573 \\ -8.1946 \end{pmatrix} ; \; x^{(4)} = \begin{pmatrix} -0.2669 \\ -0.5342 \\ -0.8021 \end{pmatrix} ; \; \lambda^{(4)} = 10.0664$$

The estimate of the dominant eigenvalue converges to the largest eigenvalue $\lambda_3 = 10$ *whereas the estimate of the dominant eigenvector converges to the corresponding eigenvector* $x_3 = (-0.2, -0.5, -0.8)^T$.

In general the sequence x_j linearly converges to x with the convergence ratio

$$\frac{|\lambda_2|}{|\lambda_1|},$$

i.e., the size of the largest eigenvalue compared with the second largest.

3.3.2 The Inverse Power Method

The inverse power method finds the smallest (in absolute value) eigenvalue and the corresponding eigenvector. Furthermore, it allows the possibility of finding any eigenvalue given its approximation. Suppose α is a close approximation of the desired eigenvalue λ_k. To compute λ_k we will apply to power method to the matrix $(A - \alpha I)^{-1}$ which has eigenvalues $(\alpha - \lambda_1)^{-1}, (\alpha - \lambda_2)^{-1}, \cdots (\alpha - \lambda_n)^{-1}$ (see the theorem below). Clearly, the largest eigenvalue of $(A - \alpha I)^{-1}$ will be the eigenvalue of A that is closest to α. If we set $\alpha = 0$ then, the method will find the smallest eigenvalue of A in absolute value.

Theorem 26. *Let* $A \in M_n$ *be nonsingular and* λ *be an eigenvalue of* A. *Then* λ^{-1} *is an eigenvalue of* A^{-1}.

Proof: Let $Ax = \lambda x$ such that $x \neq 0$. It follows that

$$x = A^{-1}(\lambda x) = \lambda A^{-1} x$$

Consequently $A^{-1}x = \lambda^{-1}x$ and thus λ^{-1} is an eigenvalue of A^{-1} $\square$

This procedure is often used to find eigenpair, when an approximate eigenpair is known. However, usefulness of this procedure is hindered by its slow convergence, which is of order

$$O(|\frac{(\lambda-\alpha)^{-1}}{(\lambda_k-\alpha)^{-1}}|)\,.$$

Here is a numerical example of the inverse power method. In the example, we do not compute A^{-1}, instead we find its LU decomposition and solve $Lz_i = x_i$ and $Uy_i = z_i$.

Example 9.

$$A = \begin{pmatrix} 5 & 2 & 2 \\ 3 & 6 & 3 \\ 6 & 6 & 9 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{3}{5} & 1 & 0 \\ \frac{6}{5} & \frac{3}{4} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 5 & 2 & 2 \\ 0 & \frac{24}{5} & \frac{9}{5} \\ 0 & 0 & \frac{21}{4} \end{pmatrix}$$

Choose $x_0 = [1\ 1\ 1]^T$ *and* $\alpha = 0$*. The table bellow summarizes the first* *iterations of the inverse power method.*

iterations	eigenvalues
1	0.3274
2	0.3684
3	0.3401
4	0.3348
5	0.3336

We can see that the eigenvalues converge to 0.3333 *which is the inverse of the smallest eigenvalue of* A, $\lambda_1 = 3.0$

4. THE FIELD OF VALUES

4.1 Introduction

The Field of Values also known as numerical range is a convex set that contains all the eigenvalues of a matrix, which is the most common reason for studying the field of values. There has been a substantial interest in researching this topic, a reflection of its importance. For this reason, this section is dedicated to the Field of Values.

Some of the definitions and important theorems that are essential for the forthcoming discussions are summarized. This section elaborates the important properties of the field of values, particularly those properties which can be used to analyze and infer information from the matrix in a way which the spectrum alone cannot provide. The field of values captures some pleasant properties of general matrices namely the properties of convexity, compactness, spectral containment, scalar multiplication, projection and positive definiteness. It describes the image of the unit sphere under the quadratic form induced by the matrix [21]. We discuss these properties while the focus of the discussion is directed towards the methods related to geometric representation of the properties of the numerical range. An important observation is the smoothness of the boundary of the field of values denoted as $\partial F(A)$. The existence of a sharp point on the boundary $\partial F(A)$ indicates an extreme point.

We discuss some of the important applications of the field of values such as the approximation of the inclusion region of all the eigenvalues and the stability analysis of dynamical systems. Thanks to the subadditivity property of the field of values more can be said about the eigenvalues of sums of matrices and the bounds of the eigenvalues of sums of matrices. Moreover, the field of values facilitates the analysis of the numerical radius and the spectral radius of sums of matrices.

4.1.1 Definitions, Important Properties and Theorems

Definition 17 (The field of values of a matrix). *The field of values of a matrix $A \in M_n$ is defined as the set*

$$F(A) = \{x^*Ax : x \in \mathbb{C}^n, x^*x = 1\}.$$

Definition 18 (Normal eigenvalue). *A point $\lambda \in \sigma(A)$ is a normal eigenvalue for the matrix $A \in M_n$ if*

- *every eigenvector corresponding to λ is orthogonal to every eigenvector of A corresponding to each eigenvalue different from λ, and*
- *the geometric multiplicity of the eigenvalue of λ (the dimension of the corresponding eigenspace of A) is equal to the algebraic multiplicity of λ (as a root of the characteristic equation of A).*

Definition 19 (Crawford number). *The Crawford number $c(A)$ of a matrix $A \in M_n$ is the smallest distance between the origin and the field of values of A; alternatively,*

$$c(A) = min\{|x| : x \in F(A)\}.$$

Definition 20 (Numerical radius)**.** *The numerical radius of a matrix $r(A)$ is defined as the quantity*

$$r(A) = max\{|x| : x \in F(A)\}.$$

There is a relationship between the infinity norm $\|\cdot\|_\infty$, Euclidian norm $\|\cdot\|_2$ and the numerical radius of a matrix $A \in M_n$. In general the relationship can be stated as: $r(A) \leq \|A\|_2$. On the other hand the relationship between the numerical radius and the spectral norm is

$$\frac{1}{2}\|A\|_2 \leq r(A) \leq \|A\|_2.$$

The furthest point from the origin in the spectrum is of great interest to us, and if we take its absolute value we achieve a very important number namely the spectral radius. The spectral radius of a matrix $\rho(A)$ is the absolute value of the point which the maximum distance from the origin is achieved. Alternatively, the spectral radius of a square matrix $A \in M_n$ is the radius of the smallest disc centred at the origin in the complex plane that circumscribes all the eigenvalues of A.

Definition 21 (Spectral radius)**.** *Spectral radius is the nonnegative real number*

$$\rho(A) = \max\{|\lambda| : \lambda \in \sigma(A)\}.$$

The spectral radius property bounds the eigenvalues of sums of matrices for example $\rho(A + B) \leq \rho\ (A) + \rho(B)$.

Proposition 5. *The average of the l_1-norm and the l_∞-norm bounds the nu-*

merical radius $r(A)$ as follows:

$$r(A) \leq \max_{1\leq i\leq n} \frac{1}{2}\Sigma_{j=1}^{n}(|a_{ij}| + |a_{ji}|)$$

$$\leq \frac{1}{2}[\max_{1\leq i\leq n} \Sigma_{j=1}^{n}|a_{ij}| + \max_{1\leq i\leq n} \Sigma_{j=1}^{n}|a_{ji}|]$$

$$\leq \frac{1}{2}(||A||_1 + ||A||_\infty).$$

Proof: See Horn on Johnson [21] pages 30–33.

If we assume the matrix A to be normal ($AA^* = A^*A$), we achieve the equality $r(A) = \rho(A)$, which is often denoted as spectraloid. The spectraloid condition is attained when either the numerical radius or the spectral radius is equal to the Euclidian norm. This implies that the bounds for the numerical radius we discussed above are also bounds for the spectral radius whenever we are dealing with normal matrices.

The following proposition combines the bounds for the numerical radius and the spectral radius.

Proposition 6. *For all $A \in M_n, \rho(A) \leq r(A) \leq \frac{1}{2}(||A||_1 + ||A||_\infty)$.*

For more details see [21, 24].

The following are some important facts about numerical range, which will be used in the subsequent parts of this chapter.

Proposition 7. *Let $A \in M_n$ then*

1. $F(A + \alpha I) = F(A) + \alpha \quad \forall\alpha \in \mathbb{C}$.

2. $F(\alpha A) = \alpha F(A) \quad \forall\alpha \in \mathbb{C}$.

3. $F(H(A)) = ReF(A)$.

4. *the spectrum of A is a subset $F(A)$.*

5. *$F(U^*AU) = F(A)$ for any U unitary.*

6. *$F(A+B) \subseteq F(A) + F(B)$.*

7. *$F(A) = Co(\sigma(A))$ if A is normal.*

Proof:

1. To prove 1 we note that this is a simple translation of the field of values of matrix A. To see that we have

$$F(A+\alpha I) = \{x^*(A+\alpha I)x : x^*x = 1, x \in \mathbb{C}^n\} = \{x^*Ax+x^*\alpha x : x^*x = 1, x \in \mathbb{C}^n\}$$

$$= \{x^*Ax+\alpha : x^*x = 1, x \in \mathbb{C}^n\} = \{x^*Ax : x^*x = 1, x \in \mathbb{C}^n\}+\alpha = F(A)+\alpha.$$

2. To show that $F(\alpha A) = \alpha F(A) \quad \forall \alpha \in \mathbb{C}$ we use the definition of the numerical range

$$\{x^*(\alpha A)x : x^*x = 1\}$$

$$= \alpha\{x^*Ax : x^*x = 1\} = \alpha F(A).$$

3. To show that $F(H(A)) = ReF(A)$ we first recall from Proposition 3 that any matrix A can be written as Hermitian part and skew Hermitian part, where the Hermitian part is $H(A) = \frac{1}{2}(A + A^*)$. We also need to consider the definition of the numerical range $x^*H(A)x : x^*x = 1, x \in \mathbb{C}^n$. We have

$$x^*(H(A))x = x^*\frac{1}{2}(A + A^*)x = \frac{1}{2}(x^*Ax + x^*A^*x)$$

$$= \frac{1}{2}(x^*Ax + (x^*Ax)^*) = \frac{1}{2}(x^*Ax + \overline{x^*Ax})$$

$$\text{Thus } F(H(A)) = Re(x^*Ax) = Re(F(A)).$$

4. Suppose x is a nonzero unit eigenvector for λ. Thus

$$x^*Ax = x^*\lambda x = \lambda x^*x = \lambda \ \forall \ x \in \mathbb{C}^n.$$

This means that $\lambda \in F(A)$. So $\sigma(A) \subseteq F(A)$.

5. Using the definition of the field of values we have

$$F(U^*AU) = \{x^*(U^*AU)x : x \in \mathbb{C}^n,\ x^*x = 1\}$$

$$x^*(U^*AU)x = (Ux)^*A(Ux) = y^*Ay \in F(A),$$

$$\text{where } y = Ux$$

$$\text{and since } y^*y = x^*U^*Ux = x^*x = 1 \text{ hence, } F(U^*AU) \subset F(A).$$

For the reverse containment we let $w = x^*Ax \in F(A)$ for all $x \in \mathbb{C}^n, x^*x = 1$, then it follows from the definition of numerical range and the unitary similarity transformation that

$$w = x^*(U^*)^{-1}U^*AUU^{-1})x = y^*U^*AUy \in F(U^*AU)$$

$$\text{where } y = U^{-1}x, \text{ hence } y^*y = 1. \text{ Thus } F(A) \subseteq F(U^*AU).$$

6. To prove the subadditivity of the field of values we let $A, B \in M_n$ and

$x \in \mathbb{C}^n : x^*x = 1$, then

$$\begin{aligned} F(A+B) &= \{x^*(A+B)x : x \in \mathbb{C}^n,\ x^*x = 1\} \\ &= \{x^*Ax + x^*Bx : x \in \mathbb{C}^n,\ x^*x = 1\} \\ &\subset \{x^*Ax : x \in \mathbb{C}^n,\ x^*x = 1\} + \{y^*By : y \in \mathbb{C}^n,\ y^*y = 1\} \\ &= F(A) + F(B). \end{aligned}$$

7. To prove that if $A \in M_n$ is normal then $F(A) = Co(\sigma(A))$, we recall that if A is normal then A can be unitarily diagonalized as we have seen in the previous chapters (see section 2.3.1). This means that with a unitary matrix U we have $U^*AU = \Lambda$ where $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$. Let $x \in \mathbb{C}^n : x^*x = 1$, then following from point 6 of this proposition (Proposition 7) we have

$$F(A) = F(U^*AU) = F(\Lambda).$$

Therefore,

$$x^*Ax = x^*\Lambda x = \sum_{i=1}^{n} \lambda_i x_i^* x_i = \sum_{i=1}^{n} |x_i|^2 \lambda_i,$$

which is a convex combination of the eigenvalues of A.

This implies that

$$F(A) = Co(\sigma(A)). \quad \square$$

A direct consequence of Proposition 7 is that the field of values of a Hermitian matrix constitute the shortest segment on the real axis containing all the eigenvalues.

Proposition 8 (Convex properties of field of values)**.** *The field of values* $F(A)$

is convex and closed.

Proof: To see that $F(A)$ is a convex subset of the complex plain see Horn and Johnson [21](pages 17–27) □

Proposition 9 (Compactness)**.** *For all $A \in M_n$ the field of values is a compact subset in the complex plane.*

Proof: A compact set implies that the set is closed and bounded. By definition the field of values of $A \in M_n$ is represented as $F(A) = \{x^*Ax : x \in \mathbb{C}^n,\ x^*x = 1\}$. Function $x \to x^*Ax$ is the continuous function representing the image of the Euclidian unit sphere. Since the Euclidian unit sphere is compact, $F(A)$ is also compact. □

To discuss about the boundary points of $F(A)$, we now turn to the geometry of the field of values. Our motivation to determine the boundary points of the field of values (denoted as $\partial F(A)$) underpins the fact that $F(A)$ is a convex and compact set.

Horn and Jonson ([21], pages 30–37) illustrated a procedure which permits the boundary points to be used as an approximation of the field of values. This is done by first rotating $F(A)$ onto the real line and producing a number of boundary points and support lines and then computing the eigenvalues and eigenvectors. First we begin by defining a sharp point.

Definition 22 (Sharp point)**.** *Let $A \in M_n$, then a point $\alpha \in F(A)$ is called a sharp point of $F(A)$ if there exist angles σ_1 and σ_2 with $0 \leq \sigma_1 \leq \sigma_2 \leq 2\pi$ such that*

$$Re\ e^{i\sigma}\alpha = \max\{Re\beta : \beta \in F(e^{i\sigma}A)\} \text{ for all } \sigma \in (\sigma_1, \sigma_2).$$

Theorem 27. *Let $A \in M_n$ and let α be a sharp point in $F(A)$ then $\alpha \in \sigma(A)$.*

Proof: See Horn and Johnson [21] pages 50–51 □.

Proposition 10. *If $\sigma(A) \subseteq \partial F(A)$, then A is normal.*

Proof: We have already noted that all the eigenvectors corresponding to normal eigenvalues are orthogonal to each other (see definition 18). This implies that A has orthonormal eigenvectors. Hence A is normal. □

Proposition 11. *If $A \in M_n$ is Hermitian then $F(A)$ is an interval in the real axis.*

Proof: To show this let λ be an eigenvalue of A, and let x be the corresponding unit eigenvector such that $Ax = \lambda x : x \neq 0$. Then $x^*Ax = \lambda x^*x$ and $\lambda = \frac{x^*Ax}{x^*x}$ which is an element of $\sigma(A)$. Note that the eigenvalues of a Hermitian matrix are necessarily real, hence $\lambda \in \mathbb{R}$. Following the spectral containment property of the field of values, $\lambda \in F(A)$ and real. □

Theorem 28. *Let $A \in M_n$ be a Hermitian with eigenvalues $\lambda_n \leq \cdots \leq \cdots \lambda_2 \leq \lambda_1$, then*

$$\frac{1}{2}(\lambda_1 - \lambda_n) \geq \max(|a_{i,j}|); \quad i \neq j.$$

Proof: See Parker [37].

4.2 The Location of The Field of Values

Identifying the location of the field of values of a matrix is a useful technique and has many applications such as numerical solution of partial differential equations, control theory, dynamical systems and solution of generalized eigenvalue problem to mention a few. The location of the field of values has been widely researched see [11, 23, 22, 24, 32, 33, 37]. In the pursuance of the topic of the location of the field of values in the complex plane we discuss about what it means the projection of the field of values on to the open right half plane (ORHP). Firstly we will identify the upper and lower bounds of the field of values using the fact that given a matrix $A \in M_n$ the field of values and the Gerschgorin's discs are inclusion sets for the spectrum of A. The relationship between the field of values and the Gerschgorin's discs can be used to pin point the set that contains the spectrum and the field of values.

To measure the relative size of the field of values as well as its location the Crawford number and the numerical radius are employed . The size of the field of values is often measured in terms of the smallest circle centered at the origin that contains the field of values. The task is to derive an iterative process that chooses successively an angle θ between 0 and 2π, which is used for rotating the matrix on to the ORHP. This permits us to calculate the largest modulus eigenvalue of the Hermitian part of the rotated matrix, which corresponds to the numerical radius of the matrix concerned. Bendixson-Hirsch theorem 30 is frequently used to provide an estimate for the upper bound of the field of values, see [32]. It provides an upper bound for the real and imaginary parts of the spectrum of a matrix. This is achieved by calculating the extreme eigenvalues of Hermitian and skew-Hermitian parts of matrices.

Theorem 29. *Let λ be the greatest eigenvalue in absolute value of matrix $A \in M_n$ and let R be the greatest sum obtained for the absolute values of the elements of a row and T be the greatest sum obtained for the absolute values of the elements of a column, then $|\lambda| \leq min(R, T)$.*

Proof: The result follows directly from the Gerschgorin Theorem (Thm. 17). □

The following theorem (from Johnson [22]) states that the projection of the field of values of $A \in M_n$ on to the real axis is achieved by taking the Hermitian part of the rotated matrix A. The following discussion on the projection of the field of values of $A \in M_n$ on to the real axis illustrates the bounds of the fields of values of a square matrix by considering the eigenvalues of the Hermitian and skew Hermitian parts of the rotated matrix A.

Lemma 1. *Let $A \in M_n$ and $x \in \mathbb{C}^n, x^*x = 1$. Then the following three conditions are equivalent.*

(i) $Re(x^*Ax) = \max_{z\in F(A)}(Re\, z)$.

(ii) $x^*H(A)x = \max_{r\in F(H(A))} r$.

(iii) $H(A)x = \lambda_{max}(H(A))x$.

Proof: For $(i) \Leftrightarrow (ii)$, we have from Proposition 7(3)

$$Re(x^*Ax) = \frac{1}{2}(x^*Ax + x^*A^*x) = x^*H(A)x.$$

For $(ii) \Leftrightarrow (iii)$, if the eigenvectors of the Hermitian matrix $H(A)$ form an orthonormal set $\{y_1, y_2, \cdots, y_n\}$ and if $H(A)y_j = \lambda_j y_j$ then x can be written as

$$x = \sum_{j=1}^{n} c_j y_j \text{ with } \sum_{j=1}^{n} \bar{c_j} c_j = 1$$

and thus

$$x^*(H(A))x = \sum_{j=1}^{n} \bar{c_j} y_j^* \lambda_j y_j c_j = \sum_{j=1}^{n} \bar{c_j} c_j \lambda_j. \ \square$$

The three equivalences can be summarized as

$$\max\{Re\alpha : \alpha \in F(A)\} = \max\{r : r \in F(H(A))\} = \lambda_{max}(H(A))\,.$$

Any matrix $A \in M_n$ can be decomposed as $A_1 + \imath A_2$ where $A_1 = \frac{1}{2}(A + A^*)$ is the Hermitian part of A and $\imath A_2 = \imath\frac{(A-A^*)}{2\imath}$ is the skew Hermitian part of A.

Proposition 12. *For any $A \in M_n$, the field of values of its Hermitian part is the projection of its field of values onto the real axis:*

$$F(A_1) = Re\, F(A)\,.$$

Proof: The field of value of A_1 contains all points of the type x^*A_1x with $x^Tx = 1$. Now we substitute, $x^*A_1x = \frac{1}{2}(x^*Ax + x^*A^*x) = \frac{1}{2}(x^*Ax + \overline{x^*Ax}) = Re\, x^*Ax$, hence the point is also contained in $Re\, F(A)$ and vice versa. □

Theorem 30. *(Bendixson-Hirsch) Let $A \in M_n$ be decomposed as $A_1 + \imath A_2$, $A_1 = \frac{1}{2}(A + A^*)$, $A_2 = \frac{(A-A^*)}{2\imath}$. Let $\lambda_1, \dots, \lambda_n, (|\lambda_1| \geq \cdots |\lambda_n|)$, $\mu_1 \geq \cdots \geq \mu_n$, $\nu_1 \geq \cdots \geq \nu_n$ be the eigenvalues of A, A_1, A_2, respectively. Then, for $k = 1, \dots, n$,*

$$\begin{aligned} \mu_n \leq & \quad Re(\lambda_k) & \leq \mu_1 \\ \nu_n \leq & \quad Im(\lambda_k) & \leq \nu_1\,, \end{aligned}$$

i.e., the eigenvalues of A lie all in the rectangle defined by intervals $[\mu_n, \mu_1]$ on the real axis and $[\nu_n, \nu_1]$ on the imaginary axis.

Proof: Let x be an eigenvector of unit length corresponding to the eigenvalue λ_k, i.e., $Ax = \lambda_k x$. Then

$$\begin{aligned} (Ax, x) = &\quad \lambda_k(x, x) &= \lambda_k \\ (A^*x, x) = &\quad \lambda_k(x, x) &= \overline{\lambda}_k\,. \end{aligned}$$

Hence

$$Re(\lambda_k) = \frac{(Ax, x) + (A^*x, x)}{2} = \left(\frac{A + A^*}{2}x, x\right) = (A_1x, x)$$

and

$$Im(\lambda_k) = \frac{(Ax, x) - (A^*x, x)}{2\imath} = (A_2x, x)\,.$$

The result now follows from the Ritz-Rayleigh theorem. □

The Bendixson-Hirsch theorem is frequently used to provide an estimate for the upper bound of the field of values. It provides an upper limit for the real and imaginary parts of the spectrum of a matrix. This is achieved by calculating the extreme eigenvalues of Hermitian and skew-Hermitian parts of matrices.

Corollary 2. *Let $A \in M_n$ be decomposed into A_1 and A_2 as in the above theorem. Then $\sigma(A) \subseteq F(A_1) + iF(A_2)$.*

Proof: Matrices A_1 and A_2 are Hermitian, and thus their field of values is equal to the convex hull of their spectrum. Hence

$$F(A_1) + iF(A_2) = \operatorname{conv}\sigma(\mathrm{A}_1) + \operatorname{conv}\sigma(\mathrm{A}_2)$$

but the latter is just the rectangle defined in the above theorem. □

Recall the Gerschgorin theorem in Section 2.3.6 that states that all eigenvalues of $A \in M_n$ are found in the union of n discs

$$\bigcup_{i=1}^{n}\{z \in \mathbb{C} : |z - a_{ii}| \leq R_r(A)\} \text{ where } R_r(A) = \sum_{j=1, j\neq i}^{n} |a_{ij}|.$$

Johnson [23] provided a Gerschgorin like estimate for the numerical radius which provides an upper bound for the spectral radius.

Theorem 31. *Let $A = [a_{ij}] \in M_n$. Then*

$$r(A) \leq \max_{1\leq i\leq n} \frac{1}{2}\{\sum_{j=1}^{n} |a_{ij}| + \sum_{j=1}^{n} |a_{ji}|\}.$$

Proof: See Johnson [23]. □

Proposition 13. *Let $A \in M_2$ with eigenvalues λ_1, λ_2. Then the field of values $F(A)$ is an elliptical disc with foci λ_1, λ_2 and minor axis of length*

$$\{trA^*A - |\lambda_1|^2 - |\lambda_2|^2\}^{\frac{1}{2}}.$$

Furthermore, if A has real eigenvalues then $F(A)$ has major axis on the real axis, which implies that $r(A) = \frac{1}{2}\rho(A + A^*)$.

Proof: See [9]. □

Proposition 14 (Positive definite indicator function)**.** *Given $A \in M_n$. Then $F(A) \subset RHP$ (right half plane) if and only if $A + A^*$ is positive definite.*

Proof: See Horn and Johnson [21, 1.2.5a]. □

4.3 Simultaneous Diagonalization by Congruence

We will now establish a link between the field of values and the solution of the Hermitian form x^*Ax.

First we recall the separating hyperplane theorem for completeness. It is used to illustrate how the field of values is rotated in the right half plane.

Theorem 32 (Separating Hyperplane Theorem)**.** *Let S_1, S_2 be two nonempty convex sets. Let S_1 be compact set and S_2 be a closed set. Assume the two sets are disjoint i.e. $S_1 \bigcap S_2 = \emptyset$. Then there exist a hyperplane separating the two sets S_1 and S_2:*

$$\exists y \in \mathbb{R}^n, y \neq 0 : \sup_{x \in S_1} y^T x < \inf_{z \in S_2} y^T z.$$

Definition 23. *We say that $A, B \in M_n$ are simultaneously diagonalizable by congruence if there exists a nonsingular $X \in M_n$ such that X^*AX, X^*BX are both diagonal.*

Proposition 15. *Let $A_1, A_2 \in M_n$ be Hermitian. Then A_1 and A_2 are simultaneously diagonalizable by congruence if and only if $A = A_1 + \imath A_2$ is congruent to a normal matrix.*

Proof: See [21, 1.7.16]. □

Theorem 33. *Assume that $A, B \in M_n$ are Hermitian. Then the following two assertions are true:*

1. *There exist $\alpha, \beta \in \mathbb{R}$ such that $\alpha A + \beta B > 0 \Longleftrightarrow 0 \notin F(A + \imath B)$.*
2. *If there are $\alpha, \beta \in \mathbb{R}$ such that $\alpha A + \beta B > 0$ then A, B are simultaneously diagonalizable by congruence.*

Proof:

1. We will first prove the $\Rightarrow$ direction. Assume $0 \in F(A+iB)$, then $\exists\ x \neq 0$ such that $x^*(A+iB)x = 0$ and thus $x^*Ax = 0$ and $x^*Bx = 0$. So for any $\alpha, \beta \in \mathbb{R}$

   $$x^*(\alpha A + \beta B)x = \alpha x^* A x + \beta x^* B x = 0.$$

   Thus $\nexists\ \alpha, \beta \in \mathbb{R}$ such that $\alpha A + \beta B > 0$.

   Now the $\Leftarrow$ direction: Assume $0 \notin F(A+iB)$. The field of values is a compact convex set, therefore there exists a separating hyperplane H that separates 0 and $F(A+iB)$. Let z be the closest point on the hyperplane to 0. In polar coordinates let $z = re^{i\theta}$. Let $\tilde{G}$ be the rotated hyperplane,

   $$\tilde{G} = e^{-i\theta} H.$$

   Then $\tilde{G}$ lies in ORHP. Since $\tilde{G}$ separates 0 and $e^{-i\theta}F(A+iB)$, also $e^{-i\theta}F(A+iB)$ lies in the ORHP. This implies that $e^{-i\theta}(A+iB)$ is positive definite. We can further write the matrix $e^{-i\theta}(A+iB)$ as $\cos\theta A - \sin\theta B + i(\cos\theta B - \sin\theta A)$. Thus taking $\alpha = \cos(\theta)$ and $\beta = \sin(\theta)$ we have $\alpha A + \beta B > 0$.

2. From the first part, we know that the assumption of positive definiteness is equivalent to $0 \notin F(A + \imath B)$. By [21, 1.7.11], $A + \imath B$ is congruent to a normal matrix and is thus simultaneously diagonalizable by the above proposition. $\square$

Corollary 3. *Let $A, B \in M_n$ be Hermitian. Then the field of values of $(A+iB)$*

does not include zero if and only if there exist an angle $\theta \in [0, 2\pi)$ such that

$$e^{i\theta} F(A + iB) \subset \text{ORHP}.$$

4.4 Calculating the Numerical Radius

The following theorem is very useful and will be used for developing an optimization problem that calculates the numerical radius of a matrix.

Theorem 34. *Let $C \in M_n$ such that $r(C)$ is the numerical radius of C. Then*

$$r(C) = \max_{\theta \in [0, 2\pi]} \lambda_{\max}(H(e^{i\theta} C))\,.$$

Proof: The proof consists of two parts. That is we will first show that

$$r(C) \geq \lambda_{\max}(H(e^{i\theta} C)), \quad \forall \theta \in [0, 2\pi]\,.$$

Then we verify the existence of an angle θ for which the equality is actually attained.

First we proof that the numerical radius is greater or equal to the largest modulus eigenvalue of the Hermitian part of the rotated matrix $(e^{i\theta} C)$. Choose $\theta \in [0, 2\pi]$. Let $x \in \mathbb{C}^n$, $\|x\|_2 = 1$ such that $x^*(H(e^{i\theta} C))x = \lambda_{max}(H(e^{i\theta} C))$.

By Lemma 1 we get

$$\begin{aligned}
\lambda_{max}(He^{i\theta}(C)) &= x^*(H(e^{i\theta}C)x \\
&\leq |x^*(H(e^{i\theta}C))x| \\
&= |\frac{1}{2}[x^*(e^{i\theta}C)^*x + x^*(e^{i\theta}C)x| \\
&= |Re(x^*e^{i\theta}Cx)| \\
&\leq |x^*e^{i\theta}Cx| \\
&= r(e^{i\theta}C) = r(C)
\end{aligned}$$

The last equation follows from the fact that $|x^*e^{i\theta}Cx| = |e^{i\theta}x^*Cx| = |x^*Cx|$, as the length of a vector is invariant w.r.t. rotation.

We now want to show that $\exists\theta \in [0, 2\pi)$ such that $\lambda_{\max}(H(e^{i\theta}C)) = r(C)$. Let $x \in \mathbb{C}^n$, $\|x\|_2 = 1$ such that $|x^*Cx| = r(C)$. Let us write the complex number x^*Cx in polar form, i.e., $x^*Cx = te^{i\theta}; t \geq 0$. Consider the rotated matrix $e^{-i\theta}C$. Then

$$x^*e^{-i\theta}Cx = e^{-i\theta}x^*Cx = e^{-i\theta}te^{i\theta} = t \in \mathbb{R}.$$

According to Lemma 1,

$$\lambda_{\max}(H(e^{-i\theta}C)) = \max_{\alpha\in F(e^{-i\theta}C)}(Re\,\alpha) = t\,,$$

where t is the numerical radius of both, C and $e^{-i\theta}C$. □

4.5 Numerical Calculations of The Field of Values, Crawford Number, and Numerical Radius using MATLAB

This section develops MATLAB programs for calculating the numerical range, numerical radius and Crawford number of a definite pair. The procedure is underpinned by an optimization problem.

MATLAB program for Calculating Numerical Radius

Given $A \in M_n, B \in M_n$ the following MATLAB program calculates numerical radius $r(C)$, where $C = (A + iB)$. The smallest circle that contains the numerical range of C is found, with the centre at $(0, 0)$ and radius $r(C)$ corresponding to the distance between the origin and the greatest eigenvalue of the pair λ_i.

The following MATLAB function computes the numerical radius r using input variables as matrix A and real k points.

```
function [r]= num_rad(A, k)
%
% numradius of A using k points
% assumes k is a positive integer
%
if nargin == 1,
end
i = sqrt(-1);
%
```

```
for l=1:k,
    th = (2*pi/k)*l;
    rotA = exp(i*th)*A;
    HrotA = (rotA+rotA’)/2;
    lmax(l) = max(eig(HrotA));
end
r = max(lmax);
```

Example 10. *This is a very simple example that uses the Kahan matrix to compute the numerical radius. A more sophisticated will yield a more efficient algorithm.*

Here is a Matlab function that constructs the Kahan matrix of dimension $n \times n$ *depending on parameter* $s = \sin(\theta)$ *and computes its numerical radius.*

```
function [K] = kahan(s,n)
% construct kahan matrix
%
A = zeros(n);
c = sqrt(1-s*s);
%
for i=1:n
    K(i,i) = s^(i-1);
    for j=i+1:n,
        K(i,j) = -c*s^(i-1);
    end
end
r=num_rad(kahan(.1,6))
```

```
r =


    1.7012
```

MATLAB program for Calculating Crawford Number

Similar to the calculation of the numerical radius of a matrix, the Crawford number of a matrix is computed using an optimization problem. The algorithm is identical to the one used for numerical radius, however, unlike the numerical radius we are interested in the smallest distance between the origin and the field of values. To achieve this, the minimum eigenvalue of the rotated matrix is required. Firstly the optimization function finds the minimum eigenvalue of the rotated matrix. Secondly it calculates the corresponding maximum angle that achieves the minimum eigenvalue.

The MATLAB function below computes the Crawford number of a matrix.

```
function [crw]= crw_simple(A, k)
%
% crawford number of A using k points
% assumes k is a positive integer
%
if nargin == 1,
    k = 100
end
i = sqrt(-1);
```

```
%

lmin = zeros(1,k);

for l=1:k,

    th = (2*pi/k)*l;

    rotA = exp(i*th)*A;

    HrotA = (rotA+rotA')/2;

    lmin(l) = min(eig(HrotA))

end

crw=max(lmin);
```

4.6 Summary

The field of values captures important properties of matrices including; convexity, compactness, spectral containment and positive definiteness to mention a few. The field of values of a matrix is essential for assertaining the set that contains its eigenvalues and the location of this set. The location of the field of values is very useful for the computations of the numerical solution of partial differential equations, control theory, dynamical systems and solution of generalized eigenvalue problem.

5. GENERALIZED EIGENVALUE PROBLEM (GEP)

5.1 Introduction

In this section the generalized eigenvalue problem $Ax = \lambda Bx$ is considered where $A, B \in M_n,\ x \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$. For the purpose of this thesis I will particularly focus on the case where both A and B are Hermitian matrices and at least one of them is positive definite.

The Generalized Eigenvalue Problems (GEP) arise naturally in many branches of science, engineering and humanities such as quantum mechanics, civil engineering, chemical engineering, and economics to mention a few. The GEP is also an important intermediary step in calculating higher polynomial eigenvalue problems such as quadratic eigenvalue problem, which will be the subject of the next chapter.

In Section 5.2 the definition of generalized eigenvalue problem and some of the mathematical properties of the generalized eigenvalue problem are discussed. The main methods available for solving GEP are explored in section 5.3. A widely used approach to solve the generalized eigenvalue problems (which will be the primary focus) is to reduce the GEP to a standard eigenvalue problem of the form $Ax = \lambda Ix$ and then apply one of the numerical methods to solve this simpler problem. This particular approach is discussed in more detail with some examples.

The generalized eigenvalue problem is a lot more complicated than the standard eigenvalue problem. The difficulty arises from the fact that the generalized eigenvalue problem my have n eigenvalues, no eigenvalues or infinite eigenvalues. GEP has n eigenvalues if and only if $rank(B) = n$. If one of the matrices is singular the number of eigenvalues of the GEP can be either infinite or zero. If A and B are Hermitian and B is positive definite we can convert the GEP to standard eigenvalue problem, the drawback is that the Hermitian property may be lost, see [12, p.394]. Fortunately, the Hermitian property can be retained if congruence transformation is used.

5.2 Mathematical Preliminaries

The generalized eigenvalue problem explores the solutions of an eigenpair of matrix pencils (A, B) by treating the problem of finding a non trivial solution to:

$$Ax = \lambda Bx.$$

The roots of the characteristic polynomials $P(\lambda) = det(A - \lambda B) = 0$ provide eigenvalues of GEP of the above matrix pencil. Since the degree of the polynomial $P(\lambda)$ is n, the number of roots of $P(\lambda)$ is also n and hence has at most n eigenvalues.

Definition 24 (Generalized Eigenvalue)**.** *The eigenvalues of A with respect to B are given by the set $\lambda(A, B)$ defined as*

$$\lambda(A, B) = \{\lambda \in \mathbb{C} : det(A - \lambda B) = 0\}.$$

Note that $\lambda(A, I) = \sigma(A)$.

Proposition 16. *Given two positive definite Hermitian matrices $A \in M_n$ and $B \in M_n$ there exist a nonzero vector x such that they satisfy the condition*

$$Ax = \lambda Bx \,, \quad \text{where } \lambda \in \mathbb{C} \text{ and } \lambda \in (A, B)\,.$$

Following our discussion on the field of values of matrices, we explore here their counterparts in matrix pencils. Indeed all the properties of the field of values of matrix can be extended to a matrix pencil which allows us to gain an insight into the location of the eigenvalues as well as their bounds. Furthermore, we can use the properties of the field of values of a pencil to check whether a pair is definite.

Definition 25 (Generalized field of values)**.** *The generalized field of values of a matrix pair $A, B \in M_n$ is a set of complex numbers*

$$W(A, B) = \{\lambda \in \mathbb{C} \mid : x^*(A - \lambda B)x = 0, \; x^*x = 1, \; x \in \mathbb{C}^n\}\,.$$

Note that for $B = I$, this definition coincides with the Definition 17.

Some authors use an equivalent definition: if A and B have a common nullspace, then $W(A, B) = \mathbb{C} \cup \{\infty\}$; otherwise

$$W(A, B) = \left\{ \frac{x^*Ax}{x^*Bx} : x \neq 0 \right\}.$$

Definition 26 (Generalized numerical radius)**.** *The numerical radius of a pencil is the radius of a circle that contains all the eigenvalues of the pair. The* *numerical radius of a pencil is defined as*

$$r(A,B) = \max\{|x^*(A+\imath B)x| : x^*x = 1,\ x \in \mathbb{C}^n\}.$$

Definition 27 (Generalized Crawford number)**.** *The Crawford number of the pair (A,B) is the smallest distance between the origin and the field of values. It is defined as*

$$c(A,B) = min\{|x^*(A+iB)x| : x \in \mathbb{C}^n,\ x^*x = 1\}\,.$$

Moreover, the Crawford number $c(A,B)$ can be interpreted as the distance from (A,B) to the nearest set of nondiagonalizable pairs. Li [35] studied this fact and presented a detailed proof. He discussed a way of computing the closest indefinite Hermitian pair by computing the Crawford number.

Proposition 17. *Let $A, B \in M_n$. Then (A,B) is congruent to $(\tilde{A},\tilde{B})$ if there exist a nonsingular matrix U such that $\tilde{A} = U^*AU$ and $\tilde{B} = U^*BU$.*

5.3 Standard Methods for Solving GEP

In this section several methods for solving GEP are treated. These methods include the QZ method, and the solution of GEP with positive definite matrices. We also discuss the method that utilizes simultaneous diagonalization by congruence. For more detail see [12].

5.3.1 QZ Method for Solving GEP

Theorem 35. *(**Generalized Schur Decomposition**) Let $A, B \in H_n$ with B nonsingular, then there exist unitary matrices Q and Z such that*

$$Q^*AZ = T = \begin{pmatrix} t_{11} & 0 & \cdots & 0 \\ 0 & t_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{nn} \end{pmatrix} \text{ and } Q^*BZ = S = \begin{pmatrix} s_{11} & 0 & \cdots & 0 \\ 0 & s_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{nn} \end{pmatrix}$$

are upper triangular with diagonal entries of t_{ii} and s_{ii} respectively; furthermore, the eigenvalues of the pencil (A, B) form the set

$$\lambda(A, B) = \{\frac{t_{ii}}{s_{ii}}; s_{ii} \neq 0\}$$

Proof: The eigenvalues satisfy

$$Ax = \lambda Bx \Longleftrightarrow (A - \lambda B)x = 0$$

$$(AB^{-1} - \lambda I)x = 0$$

Using Schur decomposition of AB^{-1}, we find $Q \in M_n$ orthogonal, such that $R_k = Q_k^*(AB_k^{-1})Q_k$ is upper triangular. Let Z_k be unitary such that $Z_k^*(B_k^{-1}Q_k) = S_k^{-1}$. Then $S_k = Q_k^*B_kZ_k$ is upper triangular, hence

$$\begin{aligned} R_kS_k &= Q_k^*(AB^{-1})Q_kQ_k^*B_kZ_k \\ &= Q_k^*(AB^{-1})B_kZ_k \\ &= Q_k^*AZ_k \end{aligned}$$

is also upper triangular. We conclude that Q^*AZ and Q^*BZ are both upper triangular (see Golub and Van Loan [12, p.377]). □

The QZ method for $Ax = \lambda Bx$ provides an effective way of solving generalized eigenvalue problem based on generalized Schur decomposition. It is a slightly modified version of QR algorithm. The QZ process unitarily transforms A and B into equivalent triangular matrices $\tilde{A}$ and $\tilde{B}$, such that $\tilde{A} = QAZ$ is triangular (Schur form) and $\tilde{B} = QBZ$ is also triangular form.

The property of positive definiteness is a very useful characteristic in analysing and solving generalized eigenvalue problem and quadratic eigenvalue problem. Given two matrices A, B which are positive definite, the generalized eigenvalue problem $Ax = \lambda Bx$ can be solved by congruence diagonalization and the eigenvalues are real and positive, see Horn and Johnson [20]. This idea will be further developed in this section.

5.3.2 Solution of GEP with Positive Definite Matrices

To solve the above generalized eigenvalue problem when either the matrix B or A is positive definite, the generalized eigenvalue problem is reduced to a standard eigenvalue problem (of the form $Cx = \lambda Ix$) by computing either AB^{-1} or $A^{-1}B$. The following is a description of these two scenarios:

1. Assume that A Hermitian (possibly singular) and B is nonsingular and positive definite. Transforming the generalized eigenvalue problem into standard eigenvalue problem is straightforward:

$$Ax = \lambda Bx \Longrightarrow B^{-1}Ax = \lambda x\,.$$

However, the drawback with this method is that the Hermitian property may be lost.

2. Now assume that A is nonsingular and positive definite and B general Hermitian (possibly singular). Using $C = A^{-1}B$, the GEP is transformed into

$$Ix = \lambda Cx\,.$$

Solving a standard eigenvalue problem

$$Cx = \mu x\,,$$

the eigenvalues of the GEP are $\lambda_i = \mu^{-1}$, $i = 1, \ldots, n$. We require that $\mu \neq 0$ otherwise the GEP eigenvalues λ will be infinite.

Example 11. *Using the above method let*

$$A = \begin{pmatrix} 9 & 2 & 1 \\ 2 & 6 & 3 \\ 1 & 3 & 11 \end{pmatrix} \text{ and } \mathrm{B} = \begin{pmatrix} 21 & 3 & 1 \\ 3 & 7 & 3 \\ 1 & 3 & 8 \end{pmatrix} \text{ then}$$

$$A^{-1}B = \begin{pmatrix} 2.4 & 0.08 & 0 \\ -0.2737 & 1.1663 & 0.1579 \\ -0.0526 & -0.0526 & 0.6842 \end{pmatrix}$$

The eigenvectors associated with $A^{-1}B$ are

$$\begin{pmatrix} 0.9747 \\ -0.2225 \\ -0.0233 \end{pmatrix} \begin{pmatrix} 0.0645 \\ -0.9928 \\ 0.1010 \end{pmatrix} \begin{pmatrix} 0.0148 \\ -0.3134 \\ 0.9495 \end{pmatrix}$$

The eigenvalues are 2.3817, 1.1680, 0.7008.

The following MATLAB routine $[AA, BB, Q, Z] = qz(A, B)$ corroborates the above result.

$$[AA, BB, Q, Z, V, W] = qz(A, B).$$

$$AA = \begin{pmatrix} 8.3215 & 4.1845 & 1.9368 \\ 0 & 5.9568 & 6.9415 \\ 0 & 0 & 9.5824 \end{pmatrix}; BB = \begin{pmatrix} 19.8197 & 7.8491 & 2.6018 \\ 0 & 6.9578 & 6.1073 \\ 0 & 0 & 6.7150 \end{pmatrix}.$$

$$Q = \begin{pmatrix} -0.9978 & 0.0654 & 0.0061 \\ -0.0640 & 0.9474 & 0.3136 \\ 0.0148 & -0.3134 & 0.9495 \end{pmatrix}; Z = \begin{pmatrix} 0.9747 & 0.2186 & 0.0475 \\ -0.2225 & 0.9694 & 0.1041 \\ -0.0233 & -0.1121 & 0.9934 \end{pmatrix}.$$

$$L = BB./AA; L = \begin{pmatrix} 2.3817 & 1.8758 & 1.3433 \\ & 1.1680 & 0.8798 \\ & & 0.7008 \end{pmatrix}$$

As expected the eigenvalues are obviously the diagonal elements of the upper-triangular matrix L, that is $2.3817, 1.1680, 0.7008$.

However, there are barriers in employing this approach, which can be categorized in three groups.

1. Firstly if B is ill-conditioned (i.e., its rows are almost linearly dependent) the eigenvalues of the computed AB^{-1} can be substantially different from the eigenvalues of (A, B).

2. Secondly AB^{-1} will not be Hermitian although A and B are. This jeopardizes the Hermitian property which greatly simplifies the computation of the eigenvalues and eigenvectors. However, the eigenvalues are still real.

3. Thirdly, if B is singular, also C is singular, where $C = A^{-1}B$, and at least one of its eigenvalues is zero. This means that λ has infinite eigenvalues corresponding to the zero eigenvalue of μ.

Example 12. *If we choose matrix B to be singular and compare the eigenvalues computed $(\lambda(A, B))$ to those of perturbed $\lambda(A, B_\epsilon)$ we find that*

$$C = AB^{-1} \Rightarrow \lambda(C) = \emptyset.$$

$$\text{Let } A = \begin{pmatrix} 5 & 1 \\ 1 & 9 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \text{ then } AB^{-1} = \emptyset.$$

This is because B is singular and the computed AB^{-1} does not result any eigenvalues.

On the other hand the eigenvalues of $\lambda(A, B_\epsilon)$ give us n eigenvalues.

$$C = AB_\epsilon^{-1} \Rightarrow \lambda(C) = \mathbb{C}$$

If we perturb B, we get $A = \begin{pmatrix} 5 & 1 \\ 1 & 9 \end{pmatrix}$ *and* $B = \begin{pmatrix} 1 & 2 \\ 2 & (4 + 1 \cdot 10^{-13}) \end{pmatrix}$,

then $\lambda(AB^{-1}) = \lambda(C) = 2.490928997439403;\quad 0.000000000000018$

which is substantially different from the result when B is singular

A possible remedy to this drawback (in the situation when A and $B \in M_n$ are Hermitian and B is positive definite) is to use congruence transformation. Section 5.3.3 discusses an algorithm which is devised to compute the eigenvalues of a matrix pencil using congruence transformation.

It is nice to have two positive definite matrices, or at least one of them to be positive definite, however, this does not always materialize in practice. Nevertheless, it suffices to find a suitable linear combination of A and B that qualifies definiteness of the pair, as we shall see.

The following is a very important theorem, which provides an insight into how to write a non definite matrix pair as definite by finding a linear combination of the pair.

Theorem 36 (Linear Combinations of Matrix Pair)**.** *A GEP with* $A, B \in H_n$ *is a definite pair if there exist* $\alpha, \beta \in \mathbb{R}$ *such that*

$$\alpha A + \beta B > 0\,.$$

The eigenvalues of $(\alpha A + \beta B, B)$ *relate to those of* (A, B)*.*

Proof: To show this let $\alpha \neq 0$, then

$$\alpha A x = \alpha \lambda B x \,.$$

Let further $\beta = \alpha\lambda$ then

$$\alpha A x + \beta B x = \alpha \lambda B x + \beta B x = (\alpha\lambda + \beta) B x \,.$$

We know that $\alpha A + \beta B > 0$, hence we can utilize Cholesky factorization to get GG^*:

$$GG^* x = (\alpha\lambda + \beta) B x \,.$$

Hence

$$x = (\alpha\lambda + \beta) G^{-1} B G^{-*} x \,,$$

where $G^{-1}BG^{-*}$ is a Hermitian matrix. This can be solved to find the eigenvalues of $G^{-1}BG^{-*}$ say $\mu_1, \mu_2, \cdots \mu_n$. And finally solve

$$(\alpha\lambda_i + \beta) = \mu_i \text{ hence } \lambda_i = \frac{\mu_i - \beta}{\alpha} \,.$$

In this situation neither of the two matrices may be positive definite, but a linear combination of the pair (A, B) is positive. □

The following example shows that it is possible to find α, β such that $\alpha A + \beta B > 0$ even if $A \ngeq 0$ and $B \ngeq 0$.

Example 13. *Let $\alpha = \beta = 1$ and $C = \alpha A + \beta B$ with*

$$A = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \textit{ and } B = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Then

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

which is clearly positive definite.

5.3.3 Solution of GEP with Congruence Transformation

When dealing with Hermitian and symmetric matrices it is natural to consider transformations that do not change the Hermitian or symmetry property. The main focus of this section is the congruence transformation, which does not affect the Hermitian or symmetry properties.

Let $X \in M_n$ be nonsingular, then

$$\underbrace{A - \lambda B}_{Hermitian\ definite} \iff \underbrace{(XAX^*) - \lambda(XBX^*)}_{Hermitian\ definite}$$

According to Definition 17 a matrix pencil (A, B) is congruent to $(\tilde{A}, \tilde{B})$ if there exist non singular matrix U such that $\tilde{A} = UAU^*$ and $\tilde{B} = UBU^*$. They do not necessarily have the same eigenvalues but their inertia is the same (see Definition 15).

Theorem 37. *If $A, B \in H_n$ are congruent and A is positive definite or positive semidefinite then B is positive definite or positive semidefinite, respectively.*

Proof: Let A be positive definite; since A is congruent to B there exist a nonsingular matrix X such that $A = X^*BX$. Let $y \in \mathbb{C}^n,\ y \neq 0$, then

$$y^*Ay = y^*X^*BXy = (Xy)^*B(Xy).$$

Let $z = Xy, z \neq 0$ then, from the positive definiteness of A, we have

$$0 < y^*Ay = (Xy)^*B(Xy) = z^*Bz$$

hence B is positive definite □

A similar approach is used when A is positive semidefinite.

We now state a well known fact on positive (semi)definite matrices:

Proposition 18 (Positive Definite Matrices)**.** *Let $A \in H_n$.*

- *A is positive definite if and only if all eigenvalues of A are positive*

- *If A is positive semidefinite and singular, then at least one of the eigenvalues is zero and the eigenvalues are nonnegative*

- *If A is indefinite then at least one of its eigenvalues is negative and at least one is positive.*

Theorem 38. *Let $A \in H_n$ be positive definite with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then there exist a unitary matrix Q such that for any $x \in \mathbb{C}$ we get*

$$x^*Ax = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

with $x = Qy$.

Proof: Let Q be a matrix formed column-wise by the eigenvectors of A. Then, since $AQ = \Lambda Q$, we get

$$x^*Ax = (Qy)^*A(Qy) = y^*Q^*AQy = y^*\Lambda y = y^* \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} y$$

and thus

$$= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2\,. \quad \square$$

Proposition 19 ([20, 4.5.8])**.** *Let $A, B \in M_n$ be Hermitian. Then there exists matrix $X \in M_n$ nonsingular such that $B = X^*AX$ if and only if A and B have the same inertia.*

When dealing with congruence transformation it is preferable to require that X to be orthonormal in order to avoid any significant element growth while conducting the congruence transformation. The fact that X is orthonormal and the unitary invariance of the norm ensures that the sequence of congruence transformations does not change the norm:

$$\|X^*AX\| = \|A\| \;\; if \;\; X^*X = I_n$$

The following simultaneous diagonalization by congruence result is a basic tool in our approach to the definite generalized eigenvalue problem. Suppose A and B are Hermitian and B is positive definite; to compute the eigenvalues and eigenvectors of generalized eigenvalue problem, firstly it has to be converted into standard eigenvalue problem of type $Cx = \lambda x$. The next theorem sheds light on this.

Theorem 39. *Let $A, B \in M_n$ be Hermitian, and B positive definite. Then there exists a nonsingular $X \in M_n$ such that*

$$X^*AX = \begin{pmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{pmatrix}, \quad X^*BX = \begin{pmatrix} \beta_1 & & & \\ & \beta_2 & & \\ & & \ddots & \\ & & & \beta_n \end{pmatrix}.$$

Moreover, $AX = \Lambda BX$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_i = \alpha_i/\beta_i$.

Proof: The simultaneous diagonalizibility is a special case of Theorem 33. The rest is obvious. □

Proposition 20. *Let $A, B \in M_n$ and let t be an eigenvalue of the generalized eigenvalue problem*

$$Ax = \lambda Bx$$

then $(t+1)$ is an eigenvalue of $(A+B)x = \lambda Bx$.

Proof: Since t is an eigenvalue of $Ax = \lambda Bx$, there exists a non-zero vector y such that $Ay = tBy$. Then

$$(A+B)y = tBy + By = (t+1)By\,. \quad \square$$

Example 14.

The following is an illustration of Matlab calculation for $eig(A, B)$ and $eig((A+B), B)$. It is clear that if t is an eigenvalue of (A, B), then $(t+1)$

is an eigenvalue of $((A+B), B)$.

$$A = \begin{pmatrix} 0.8147 & 0.9134 & 0.2785 \\ 0.9058 & 0.6324 & 0.5469 \\ 0.1270 & 0.0975 & 0.9575 \end{pmatrix} B = \begin{pmatrix} 1.8673 & 0.6765 & 1.8324 \\ 0.6765 & 0.4383 & 0.9276 \\ 1.8324 & 0.9276 & 2.4211 \end{pmatrix}$$

$$eig(A, B) = 10.2381, -0.5455, 0.5225$$

$$eig((A+B), B) = 11.2381, 0.4545, 1.5225$$

5.4 Summary

The solution of GEP is more difficult compared to SEP. There are several methods for solving the GEP. The QZ method which is based on generalized Schur decomposition unitarily transforms A and B into triangular forms. This greatly simplifies the computation of the eigenvalues of the pencil, which forms the set $\{\lambda(A, B) = \frac{t_{ii}}{s_{ii}},\ s_{ii} \neq 0\}$.

The GEP is reduced to SEP when either A or B is positive definite by computing either $A^{-1}B$ or AB^{-1}. In the event when neither A nor B is positive definite we find $\alpha,\ \beta \in \mathbb{R}$ such that a linear combination $\alpha A + \beta B$ is positive definite.

One may want to preserve the Hermitian property of the pair. In this regard the natural choice is congruence transformation which does not affect Hermitian property of the pair.

6. QUADRATIC EIGENVALUE PROBLEM

6.1 Introduction

The quadratic eigenvalue problem is a very important nonlinear eigenvalue problem. It has wide applications in many engineering and scientific areas such as structural dynamics and nonlinear vibration theory. See Meerbergen and Tisseur [42] for an in depth discussion. Most QEP computations require only a few eigenvalues, normally, either the smallest, or the largest eigenvalues in modulus suffice. Guo et al [14] introduced a numerical algorithm which finds the few smallest eigenvalues and the few largest negative eigenvalues. A significant research is directed to the solution of large scale quadratic eigenvalue problem. For example, Ye [45] applied Arnoldi algorithm based on Krylov subspaces generated by a single matrix $A^{-1}B$. Shift and invert transformation has been widely used for both standard eigenvalue problem and quadratic eigenvalue problem. It is well known for its ability to accelerate the convergence of Krylov subspace method. Lin and Bao [28] used block second order Arnoldi procedure and block second order biorthogonalisation procedure (a procedure which generates the biorthonormal basis of the second order right and left Krylov subspaces) to produce biorthonormal basis. Consequently this has been used to reduce the size of large scale quadratic eigenvalue problem without compromising the necessary properties of the QEP.

The most usual way of solving QEP is to convert it into linear form. There has been a vast literature on the linearisation techniques some of which is discussed in this chapter, see [19], [1], [29], [16].

The main focus here is the hyperbolic quadratic eigenvalue problem in which the eigenvalues are all real, similar to the definite generalized eigenvalue problem and the standard Hermitian eigenvalue problem which also have real eigenvalues as shown by [16], [18].

Definition 28. *Given $A, B, C \in M_n$, the quadratic eigenvalue problem finds $\lambda \in \mathbb{C}$ and $x \in \mathbb{C}^n (x \neq 0)$ such that*

$$Q(\lambda)x = 0, \text{ where } Q(\lambda) = \lambda^2 A + \lambda B + C.$$

If λ and x solve the QEP, then we call λ a quadratic eigenvalue and x a quadratic eigenvector for A, B, C. Also this can be interpreted as a matrix polynomial of degree 2.

The spectrum of $Q(\lambda)$ is defined as the set of all eigenvalues

$$\sigma(Q(\lambda)) = \{\lambda \in \mathbb{C} : det\, Q(\lambda) = 0\}$$

A quadratic eigenvalue problem is said to be regular if the roots of its characteristic equation is not identically zero for λ. Since $det(Q(\lambda))$ is a polynomial of degree $2n$, then $|\sigma(Q(\lambda))| \leq 2n$.

Throughout this chapter A is assumed to be nonsingular. If $Q(\lambda)x = 0,\ x \neq 0$, then we can form Rayleigh quotient for QEP. Given x, $x^*x = 1$, $x \neq 0$, as a right eigenvector for QEP such that

$$Q(\lambda)x = (\lambda^2 A + \lambda B + C)x = 0,$$

we obtain

$$x^* Q(\lambda)x = \lambda^2 x^* Ax + \lambda x^* Bx + x^* Cx = 0;$$

and

$$\lambda = \frac{-x^* Bx \pm \sqrt{(x^* Bx)^2 - 4(x^* Ax)(x^* Cx)}}{2x^* Ax}.$$

At least one of the solutions to the above quantity is an eigenvalue, while the other might not be.

Definition 29. *Let A, B, C be Hermitian matrices and $A > 0$. Then the quadratic eigenvalue problem $Q(\lambda)$ in Definition 28 is said to be hyperbolic if the quantity $(x^* Bx) - 4(x^* Ax)(x^* Cx)$ is positive for all $x \neq 0$. That is,*

$$(x^* Bx)^2 > 4(x^* Ax)(x^* Cx) \; \forall x \in \mathbb{C}^n, x \neq 0.$$

For hyperbolic QEP all $2n$ eigenvalues are real. Moreover, there is a gap between the n largest eigenvalues (often called the principal eigenvalues) and the n smallest eigenvalues (known as secondary eigenvalues) with n linearly independent eigenvectors associated with each of the primary and the secondary eigenvalues; see [13], [2].

Definition 30. *A hyperbolic quadratic eigenvalue problem is overdamped if A and B are positive definite and C positive semidefinite.*

Definition 31. *A QEP is elliptic if A is Hermitian positive definite, B and C*

are Hermitian and $(x^*Bx) - 4(x^*Ax)(x^*Cx)$ *is negative for all non zero* $x \in \mathbb{C}^n$. *That is*

$$(x^*Bx)^2 < 4(x^*Ax)(x^*Cx) \ \forall x \in \mathbb{C}^n, x \neq 0.$$

The following table from [42] provides an insight into the different matrix properties and the corresponding properties of their eigenvalue and eigenvectors.

	Matrix properties	Eigenvalue properties	Eigenvector properties
1	A singular	$2n$ finite eigenvalues	
2	A singular	Finite and infinite eigenvalues	
3	A, B, C real	Eigenvalues are real or come in pairs $(\lambda, \bar{\lambda})$	If x is a right eigenvector of λ then $\bar{x}$ is a right eigenvector of $\bar{\lambda}$
4	A, B, C Hermitian	Eigenvalues are real or come in pairs $(\lambda, \bar{\lambda})$	If x is a right eigenvector of λ then $\bar{x}$ is a right eigenvector of $\bar{\lambda}$
5	A Hermitian positive definite, B, C Hermitian positive semidefinite	$Re(\lambda) \leq 0$	
6	A, B symmetric positive definite, C symmetric $\gamma(A, B, C) > 0$	λs are real and negative, gap between n largest and n smallest eigenvalues	n linearly independent eigenvectors associated with the n largest (n smallest) eigenvalues
7	A, B Hermitian, A positive definite, $B = -B^*$	Eigenvalues are purely imaginary or come in pairs	If x is a right eigenvector of λ then x is a right eigenvector of $-\bar{\lambda}$
8	A, C real symmetric and positive definite, $C = -C^T$	Eigenvalues are purely imaginary	

Proposition 21 (Hyperbolic QEP). *A QEP with* $A, B,$ *and* $C \in M_n$ *Hermitian and* A *positive definite is hyperbolic if and only if* $Q(\mu)$ *is negative definite for some* $\mu \in \mathbb{R}$.

Proof: We know that $Q(\mu) < 0$ is equivalent to $\mu^2 A + \mu B + C < 0$, hence

$$\begin{aligned}
\mu^2 A + C &< -\mu B \\
\mu^2 \underbrace{x^* A x}_{\in \mathbb{R}} + \underbrace{x^* C x}_{\in \mathbb{R}} &< -\mu x^* B x, \ \forall x \neq 0 \\
\text{and since } \ \tfrac{a+b}{2} \geq \sqrt{ab}, \ \text{ we have} & \\
2\sqrt{\mu^2 (x^* A x)(x^* C x)} &< -\mu(x^* B x) \\
0 < 2|\mu|\sqrt{(x^* A x)(x^* C x)} &< -\mu(x^* B x) \\
\Rightarrow 4\mu^2 (x^* A x)(x^* C x) &< \mu^2 (x^* B x)^2 \\
4(x^* A x)(x^* C x) &< (x^* B x)^2 \\
(x^* B x)^2 - 4(x^* A x)(x^* C x) &> 0\,. \quad \square
\end{aligned}$$

Since the quadratic eigenvalue problem $Q(\mu)$ is a nonlinear continuous function of μ with coefficient matrices A, B and C we can find the derivative $Q'(\mu)$ by differentiating the entries with respect to μ:

$$\begin{aligned}
Q(\mu) &= \mu^2 A + \mu B + C \\
Q'(\mu) &= \lim_{h \to 0} \frac{Q(\mu + h) - Q(h)}{h} \\
&= \lim_{h \to 0} \frac{[(\mu + h)^2 - \mu^2]A + [(\mu + h) - \mu]B}{h} \\
&= \lim_{h \to 0} \frac{h[(2\mu + h)A] + Bh}{h} \\
&= \lim_{h \to 0} (2\mu A + hA + B) \\
&= 2\mu A + B.
\end{aligned}$$

If for a given μ we have $Q'(\mu) > 0$ that is $2\mu A + B > 0$, then $Q(\mu)$ is increasing at μ. That is $\exists\epsilon > 0$ such that

$$\forall \eta \in (0, \epsilon),\ \ Q(\mu + \eta) > Q(\mu).$$

Furthermore,

$$Q'(\mu) > 0 \Leftrightarrow 2\mu A + B > 0$$

$$\Leftrightarrow 2\mu A > -B \Leftrightarrow \mu I > -\frac{1}{2}A^{-1}B.$$

$$\Leftrightarrow \mu > \frac{1}{2}\mu_{max}(A^{-1}B) = \frac{1}{2}\mu_{max}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}).$$

Following from Proposition 21, if $A > 0$ and $\exists\ \mu$ such that $Q(\mu) < 0$ then $Q(\mu)$ is hyperbolic.

A quadratic eigenvalue $Q(\lambda)$ has real eigenvalues if it is hyperbolic and complex eigenvalues if it is elliptic. To illustrate this fact the following example from [16] is drawn using the MATLAB routing *polyeig*.

The following shows the fact that the eigenvalues of elliptic quadratic eigenvalues are always imaginary.

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix};\ B = \begin{pmatrix} 3.5 & 0 & 0 \\ 0 & 7.5 & 0 \\ 0 & 0 & 5 \end{pmatrix};\ C = \begin{pmatrix} 3.5 & 1 & 0 \\ 1 & 8 & 1 \\ 0 & 1 & 4 \end{pmatrix}.$$

$$[e] = polyeig(A, B, C).\ e = \begin{matrix} -0.2659 + 0.7027i & -0.2659 - 0.7027i & -0.6624 + 0.7494i \\ -0.6624 - 0.7494i & -0.4796 + 0.4203i & -0.4796 - 0.4203i \end{matrix}$$

The following shows the fact that the eigenvalues of hyperbolic quadratic eigenvalue problem are always real.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}; \; B = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 30 & 0 \\ 0 & 0 & 20 \end{pmatrix}; \; C = \begin{pmatrix} -1 & 2 & 0 \\ 2 & 8 & 2 \\ 0 & 2 & 0 \end{pmatrix}$$

$[e] = polyeig(A, B, C);\; e = \;\; 64.9815, \;\; 5.4489, \;\; -2.9371, \;\; -0.1403, \;\; -0.1028, \;\; -0.2502$

There are two main methods for solving QEP which are featured in the QEP literature. The two methods are the linearization approach and the factorization approach.

The linearization method which is the subject of the next section attempts to find the $2n$ eigenvalues of QEP by solving generalized eigenvalue problem (GEP) for matrices of dimensions twice the size of the original QEP.

The factorization method solves the closely related matrix equation $Q(X) = AX^2+BX+C = 0$, where $Q(X)$ has at least two solutions (often called solvents). Higham and Kim [17] showed using Bernoulli iteration that if the gap between the primary and the secondary eigenvalues is large, the factorization method is more efficient for computing the eigenvalues of over damped QEPs.

6.2 Linearization Techniques

The linear eigenvalue problems such as Standard Eigenvalue Problems (SEP) and Generalized eigenvalue Problems (GEP) enjoy a wealth of techniques and numerical algorithms. For example, using the methods such as Schur form, generalized Schur form and simultaneous diagonalization of definite GEP, which are mentioned in the previous chapters (see sections 2 and 5.3.3).

The methods used for solving QEP are a lot more difficult than those used for their linear counterparts. Fortunately, some of the techniques used for solving linear eigenvalue problems can be adapted for the solution of QEP. In the process, the quadratic eigenvalue problem is transformed into a linear eigenvalue problem such as GEP. Once linearized, we can employ the rich techniques and algorithms such as QZ, Lanczos and Krylov algorithms. Furthermore, the process of transforming a nonlinear QEP into a linear GEP does not affect the eigenvalues.

We frequently confront some of the coefficient matrices of $Q(\lambda)$ having nice properties. We would want to preserve these properties and exploit them.

Definition 32 (Linearization of QEP)**.** *Let $Q(\lambda)$ be a matrix polynomial of degree 2. A linear generalized eigenvalue problem $L(\lambda) = \lambda X + Y$ with twice the size of $Q(\lambda)$ is called linearization of $Q(\lambda)$ if and only if there exist constant λ-matrices with nonzero determinant $E(\lambda), F(\lambda)$ such that*

$$\begin{bmatrix} Q(\lambda) & 0 \\ 0 & I_n \end{bmatrix} = E(\lambda)L(\lambda)F(\lambda).$$

Clearly, $Q(\lambda)$ and $L(\lambda)$ have the same spectrum.

To linearize a QEP, we find $X, Y \in M_n$ such that $det(\lambda X + Y) = 0 \Leftrightarrow det(Q(\lambda)) = 0$. There are an infinite number of ways that the quadratic eigenvalue problem can be linearized. The linearization method may influence the accuracy and the stability of the computed solution. Therefore, it is important to take into account the accuracy of the linearization method used. Likewise, it is important to check whether the linearization method preserves the favourable properties of the QEP (often called structure preserving transformation).

The time domain version of the QEP in definition (28) is

$$A\ddot{u} + B\dot{u} + C = 0\,.$$

This equation can be reduced to first order system and subsequently solved using one of the methods used for linear eigenvalue problem.

The following are some of the most prominent standard linearization techniques.

1. Method 1: This provides a system based on the above second order differential equation augmented with a nonsingular matrix E.

   $$\text{let } \dot{u} = v \text{ then } \begin{pmatrix} E & 0 \\ 0 & A \end{pmatrix} \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 & E \\ -C & -B \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

   In fact any nonsingular matrix may be chosen for augmentation; however, the best choice for augmentation matrix E is undoubtedly the identity matrix I. This is the best choice because the identity matrix is nonsingular, positive definite, diagonal, compact and it is always easy to deal with it computationally. Here is the above method 1 augmented with the identity matrix.

   $\{I\dot{u} = I\dot{u}$ and $A\ddot{u} = -Cu - B\dot{u}\}$

   $$\begin{pmatrix} I & 0 \\ 0 & A \end{pmatrix} \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 & I \\ -C & -B \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

2. Method 2: We could derive symmetric formulation by multiplying the

first row of method 1 by C and the second row by -1

$$\begin{pmatrix} C & 0 \\ 0 & -A \end{pmatrix} \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 & C \\ C & B \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

The advantage of this method is that we can get a Hermitian pencil if matrix B is Hermitian.

3. Method 3: If the rows of the state vectors in method 1 are swaped provides the following system.

$$\begin{pmatrix} 0 & I \\ A & B \end{pmatrix} \begin{pmatrix} \dot{v} \\ \dot{u} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & -C \end{pmatrix} \begin{pmatrix} v \\ u \end{pmatrix}$$

4. Method 4: If matrix B is Hermitian we could derive a Hermitian system from above by multiplying first row of the above equation by A.

$$\begin{pmatrix} 0 & A \\ A & B \end{pmatrix} \begin{pmatrix} \dot{v} \\ \dot{u} \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & -C \end{pmatrix} \begin{pmatrix} v \\ u \end{pmatrix}$$

As expected, the result is a linear GEP which is also Hermitian.

With this motivation, we are now coming to linearization techniques for QEP. Here are two possible linearizations that preserve the Hermitian property.

Theorem 40. *Let A, B and $C \in M_n$ be Hermitian and A positive definite. Furthermore, let C be nonsingular. Then*

1.

$$\begin{pmatrix} 0 & C \\ -C & -B \end{pmatrix} z = \lambda \begin{pmatrix} C & 0 \\ 0 & A \end{pmatrix} z$$

is a linearization of $Q(\lambda)$.

2.

$$\begin{pmatrix} -C & 0 \\ 0 & A \end{pmatrix} z = \lambda \begin{pmatrix} B & A \\ A & 0 \end{pmatrix} z$$

is a linearization of $Q(\lambda)$.

Proof: We need to find matrices $E(\lambda)$ and $F(\lambda)$ satisfying the definition of linearization. Here they are:

1.

$$E(\lambda) = \begin{pmatrix} -(B+\lambda A)C^{-1} & -I \\ C^{-1} & 0 \end{pmatrix}, \quad F(\lambda) = \begin{pmatrix} I & 0 \\ \lambda I & I \end{pmatrix},$$

2.

$$E(\lambda) = \begin{pmatrix} -I & -\lambda I \\ 0 & A^{-1} \end{pmatrix}, \quad F(\lambda) = \begin{pmatrix} I & 0 \\ \lambda I & I \end{pmatrix}. \quad \square$$

The above linearized GEP may be solved using the methods such as QZ or other iterative algorithms, or even in some special circumstances can be reduced to a standard eigenvalue problem (SEP). For example if Y is nonsingular it is possible to calculate the few minimum and the few maximum eigenvalues using Lanczos method and Arnoldi method.

Here is a numerical example to illustrate this. I use an example which was used by Hachez and Van Dooren [15]

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}; \; B = \begin{pmatrix} 3.5 & 0 & 0 \\ 0 & 7.5 & 0 \\ 0 & 0 & 5 \end{pmatrix}; \; C = \begin{pmatrix} 3.5 & 1 & 0 \\ 1 & 8 & 1 \\ 0 & 1 & 4 \end{pmatrix}$$

$$X = [C \text{ zeros}(3); \text{zeros}(3)\ A] = \begin{pmatrix} 3.5 & 1 & 0 & 0 & 0 & 0 \\ 1 & 8 & 1 & 0 & 0 & 0 \\ 0 & 1 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix}$$

$$Y = \text{[-C -B; zeros(3) -C]} = \begin{pmatrix} -3.5 & -1 & 0 & -3.5 & 0 & 0 \\ -1 & -8 & -1 & 0 & -7.5 & 0 \\ 0 & -1 & -4 & 0 & 0 & -5 \\ 0 & 0 & 0 & -3.5 & -1 & 0 \\ 0 & 0 & 0 & -1 & -8 & -1 \\ 0 & 0 & 0 & 0 & -1 & -4 \end{pmatrix}$$

$$L = inv(X) * Y = \begin{pmatrix} -1 & 0 & 0 & -1.0383 & 0.2871 & -0.0478 \\ 0 & -1 & 0 & 0.1340 & -1.0048 & 0.1675 \\ 0 & 0 & 1 & -0.0335 & 0.2512 & -1.2919 \\ 0 & 0 & 0 & -1.7500 & -0.5000 & 0 \\ 0 & 0 & 0 & -0.3333 & -2.6667 & -0.3333 \\ 0 & 0 & 0 & 0 & -0.2500 & -1 \end{pmatrix}$$

Using the Matlab routine for QZ method of $\lambda = eig(X, Y)$ we get the eigenvalues of the pencil $Y = \lambda X$

$$\lambda = \begin{pmatrix} -0.6525 + 0.7470i, & -0.6525 - 0.7470i, & -0.4424 + 0.4531i, \\ -0.4424 - 0.4531i, & -0.5726 + 0.5042i, & -0.5726 - 0.5042i \end{pmatrix}$$

We compare the above result with the answer achieved from the Matlab routine for polynomial eigenvalue $e = polyeig(A, B, C)$

$$\lambda = \begin{pmatrix} -0.6525 + 0.7470i, & -0.6525 - 0.7470i, & -0.5726 + 0.5042i, \\ -0.5726 - 0.5042i, & -0.4424 + 0.4531i, & -0.4424 - 0.4531i \end{pmatrix}$$

As expected the two results are identical.

The GEP may be sensitive to perturbations as small changes in the elements of the matrix pair may result the condition number of the GEP to increase significantly.

6.2.1 Factorization Method

Although the linearization technique is the standard method of solving QEP, the factorization method is more efficient for the solution of hyperbolic quadratic eigenvalue problems ($HQEP$), see [14], [31]. There is rich literature for the solution of lambda matrices by the way of factorization (see [4]).

The quadratic matrix equation ($AX^2 + BX + C = 0,$ where $A, B, C \in M_n$) and the quadratic eigenvalue problem are related. The relationship stems from the theory of λ matrices. However, the solution of the quadratic matrix equation is quite challenging for it can have a finite positive number of solutions (called solvents), no solutions or infinitely many solutions.

The formula used for quadratic scalar function does not generally satisfy the quadratic matrix equation.

We use the following generalized Bézout theorem to motivate the factorization method.

Theorem 41 (Generalized Bézout)**.** *Let $Q(\lambda) = \lambda^2 A + \lambda B + C$. Then $Q(\lambda)$ can be factorized as*

$$Q(\lambda) = (\lambda A + AX + B)(\lambda I - X)$$

if and only if X is a solution of the corresponding quadratic matrix equation

$$AX^2 + BX + C = 0.$$

Proof: Follows from the following system of equalities:

$$\begin{aligned} Q(\lambda) - Q(X) &= (\lambda^2 A - AX^2 + \lambda B - BX) \\ &= A(\lambda^2 I - X^2) + B(\lambda I - X) \\ &= (\lambda A + AX + B)(\lambda I - X)\,. \qquad \Box \end{aligned}$$

The above quantity can be solved explicitly if $A = I$, B commutes with C and $B^2 - 4C$ has a square root then $X = \frac{1}{2}B + \frac{1}{2}(B^2 - 4C)^{\frac{1}{2}}$, see Guo and Lancaster [13].

6.3 Hyperbolic QEP

The definite *GEP*, *HQEP* and Hermitian eigenvalue problem share many desirable properties. One of the important similarities they share is (as mentioned in definition 29 and proposition 29) the fact that their eigenvalues are always real. This section investigates the possibility of adopting the techniques used for solving Hermitian eigenvalue problem and definite GEP for solving $HQEP$ and whether appropriate transformation can be employed to calculate the eigenvalues of $HQEP$ without compromising the hyperbolicity property. Also some of the key characteristics of $HQEP$ are discussed.

6.3.1 Testing QEP for Hyperbolicity and Ellipticity

It is important to know whether the QEP in hand is Hyperbolic or elliptic. There is a rich research interests directed towards finding methods for testing QEP for hyperbolicity and ellipticity, see [15], [16].

Marcus [32] showed that testing for hyperbolicity amounts to testing the linearized Hermitian pencil for definiteness. The result in Section 5.3 can be used to achieve this. Similarly, we have shown how to check whether the Crawford number of a Hermitian matrix is positive by solving the following constrained one dimensional global optimization problem:

$$\text{minimize } \lambda \text{ subject to } L(\lambda) = 0,$$

where $L(\lambda)$ is a linearization of $Q(\lambda)$.

A similar approach has been adopted by Higham et al. [16] using bisection and level set method. They showed that to test for hyperbolicity it suffices to solve the one-dimensional global optimization problem. However, the bisection method is quite slow, which is a major weakness.

They have also proposed more appropriate and efficient algorithm (Algorithm 2.3 copied here for completeness) for testing a QEP for hyperbolicity or ellipticity.

Given the hyperbolic QEP

$$Q(\lambda) = \lambda^2 A + \lambda B + C$$

we linearize it to get the Hermitian GEP, $L(\lambda) = Xv - \lambda Y v$. Given the two Hermitian matrices X and Y with $X + iY$ non-singular and positive definite, we want to compute the quantity

$$\omega(X,Y) = \max\{\frac{1}{2}\lambda_{min}(e^{-\imath\theta}(X + \imath Y) + e^{\imath\theta}(\overline{X + \imath Y}))\}.$$

The following is Algorithm 2.3 from Higham et al [16]. It refines the bracket to an interval of width at most *tol* containing the Crawford number $\omega(X,Y)$.

1. Set a bracket $[a,b]$ for $\omega(X,Y)$ with

$$a = f(\theta) = \frac{1}{2}\lambda_{min}(e^{-\imath\theta}(X + \imath Y) + e^{\imath\theta}(\overline{X + \imath Y})), \text{ for some } \theta \in [0, 2\pi),$$

$$b = \sigma\left\{\begin{bmatrix} A \\ B \end{bmatrix}\right\}.$$

2. while $b - a > tol$

    $\xi = (a+b)/2$

    Let $Q(z) = C - 2\xi z I + z^2 C^*$, $C = X + iY$. Compute eigenvalues z_j of $Q(z)$.

    If $\lambda_{min}(e^{-\imath\theta}(X + \imath Y) + e^{\imath\theta}(X + \imath Y)) = \xi$ for some eigenvalue $z_j = e^{\imath\theta}$ of Q on the unit circle, then

        $a = \xi$

```
        else
            $b = \xi$
            if $b \leq 0$; return with $a = b = 0$, end        $(\omega(X, Y) = 0)$
        end
    end
```

The algorithm terminates when it reaches a negative b, which is an indicative of the pair being non-definite. As our aim is to test whether (X, Y) is definite, we can terminate the algorithm as soon as the lower bound a is positive.

Higham et al. [16] suggested an alternative algorithm for testing the definiteness of the pair, hence hyperbolicity. It solves just one quadratic eigenvalue problem, thus making it is more efficient than the previous algorithm. This algorithm is copied here for completeness.

The following algorithm is Algorithm 2.4 from Higham et al [16]. Given a Hermitian pair A, B with $A+iB$ non-singular this algorithm determines whether or not the pair is definite.

1. Compute the eigenvalues of $Q(z)$ (with $\xi = 0$).

2. If there are $2n$ eigenvalues of unit modulus

    Compute the derivatives of the above definite pair. We know from Proposition 21 that the derivative of $Q(\lambda) = Q'(\lambda)$ can be achieved since the quadratic eigenvalue problem $Q(\lambda)$ is differentiable with coefficients matrices A, B and C. Hence the derivative of a simple zero eigenvalue of $A_\theta - \xi I$ with

```
        normalized eigenvector $v$ is given by

        $Q'(z) = \frac{\partial}{\partial\theta}\lambda_i(A_\theta - \xi I)|_{\theta_j} = v^* \frac{\partial}{\partial\theta}(A_\theta - \xi I)|_{\theta_j} v = v^* B_{\theta_j} v$

        with $\xi = 0$.
    If there are $n$ consecutive strictly increasing and $n$ consecutive
        strictly decreasing zero crossings
            The pair is definite; return
    end
3. end
4. The pair is not definite.
```

The weakness of Algorithm 2.3 is the fact that it is too expensive since it tries to compute all the $2n$ eigenvalues of the quadratic eigenvalues problem, making it more expensive and slower compared to Algorithm 2.4. However, it has one advantage over Algorithm 2.4 namely it produces a bracket for the Crawford number $\gamma(A, B)$, which shrinks to zero. Whereas the Algorithm 2.4 produces only a monotonically increasing lower bound, see [16].

The following result shows a very efficient procedure which can be used to test the linearized $L(\lambda)$ for hyperbolicity.

Theorem 42. *A QEP with A, B, C Hermitian and A positive definite is hyperbolic if and only if the pair (A_1, A_2) is definite, where*

$$A_1 = \begin{pmatrix} -C & 0 \\ 0 & A \end{pmatrix}, \; A_2 = -\begin{pmatrix} B & A \\ A & 0 \end{pmatrix}.$$

Proof: Assume (A_1, A_2) is a definite pair, and there exist $\alpha, \beta \in \mathbb{R}$ such that

$$\alpha A_1 + \beta A_2 > 0.$$

That is

$$\alpha \begin{pmatrix} -C & 0 \\ 0 & A \end{pmatrix} - \beta \begin{pmatrix} B & A \\ A & 0 \end{pmatrix} > 0 \qquad (\star)$$

We must show that $Q(\lambda) = \lambda^2 A + \lambda B + C$ is hyperbolic. Since $A > 0$, examination of the $(2, 2)$ block implies $\alpha > 0$. Consequently there are three cases to consider: $\beta = 0$, $\beta < 0$ and $\beta > 0$.

First suppose $\beta = 0$. In this case

$$\alpha \begin{pmatrix} -C & 0 \\ 0 & A \end{pmatrix} > 0 \quad \Rightarrow -\alpha C > 0 \text{ and } \alpha A > 0\,.$$

We know that $\alpha > 0$, thus $C < 0$. Hence $(x^*Bx)^2 > (x^*Ax)(x^*Cx)$ for all $x \neq 0$ and thus $Q(\lambda)$ is hyperbolic.

When $\beta \neq 0$, we are left with the two choices $\beta > 0$ or $\beta < 0$. We will first address the case when $\beta < 0$. From $(\star)$ in the previous page we have

$$\frac{\alpha}{\beta} \begin{pmatrix} -C & 0 \\ 0 & A \end{pmatrix} - \begin{pmatrix} B & A \\ A & 0 \end{pmatrix} > 0$$

$$\Leftrightarrow \begin{pmatrix} -\frac{\alpha}{\beta}C - B & -A \\ -A & \frac{\alpha}{\beta}A \end{pmatrix} > 0$$

The Schur complement argument of the above is as follows:

$$(-\frac{\alpha}{\beta}C - B) - A(\frac{\alpha}{\beta}A)^{-1}A > 0$$

$$(-\frac{\alpha}{\beta}C - B) - \frac{\beta}{\alpha}A > 0$$

$$Let\ \lambda = \frac{\beta}{\alpha}$$

$$-B - \lambda^{-1}C - \lambda A > 0$$

$$\Rightarrow \lambda^2 A + \lambda B + C < 0$$

Thus we have found that for $\lambda = \frac{\beta}{\alpha}$, $Q(\lambda) < 0$.

For the only if part we have: By Proposition 21, if $\lambda = 0$ we can always choose $\tilde{\lambda} = \epsilon$ such that $\epsilon > 0$ then

$$Q(\tilde{\lambda}) = Q(\lambda) + \epsilon^2 A + \epsilon B + C > 0\,.$$

Assume $\lambda \neq 0$. There are two cases to consider, $\lambda > 0$, and $\lambda < 0$. The case when $\lambda < 0$ is slightly more complicated, so we will consider only this case.

There is λ such that

$$\lambda^2 A + \lambda B + C < 0.$$

Let $\alpha = 1,\ \beta = \ -\lambda$:

$$\beta^2 A - \beta B + C < 0\,.$$

Dividing by β we have

$$\beta A - B + \beta^{-1}C > 0$$

$$\Rightarrow \begin{pmatrix} B-\beta^{-1}C & A \\ A & \beta^{-1}A \end{pmatrix} > 0$$

$$\Rightarrow \begin{pmatrix} \beta B-C & \beta A \\ \beta A & A \end{pmatrix} > 0$$

$$\beta \begin{pmatrix} B & A \\ A & 0 \end{pmatrix} + 1 \begin{pmatrix} -C & 0 \\ 0 & A \end{pmatrix} > 0$$

as required. □

Also of interest is finding the degree we need to perturb the coefficient matrices of hyperbolic and elliptic quadratic eigenvalue problems so that these nice properties are lost. This entails to finding the distance problems to the nearest non-hyperbolic or non-elliptic quadratic eigenvalue problems. Clearly, both properties are lost when the coefficient matrix A of the quadratic term is perturbed to lose definiteness. The minimum perturbation which allows the loss of these properties is achieved when A is perturbed by a matrix ΔA whose $\|.\|_2$-norm is equal to $\lambda_{min}(A)$.

We introduce the following matrix

$$W(x,A,B,C) = \begin{pmatrix} 2x^*Ax & x^*Bx \\ x^*Bx & 2x^*Cx \end{pmatrix}$$

Hachez and VanDooren [15] used trigonometric matrix polynomial to obtain optimal perturbations

$$P(\omega) = \begin{pmatrix} \sin(\omega) & \cos(\omega) \end{pmatrix} \begin{pmatrix} A & \frac{B}{2} \\ \frac{B}{2} & A \end{pmatrix} \begin{pmatrix} \sin(\omega) \\ \cos(\omega) \end{pmatrix}$$

$$P(\omega) = \sin^2(\omega)A + \sin(\omega)\cos(\omega)\frac{B}{2} + \sin(\omega)\cos(\omega)\frac{B}{2} + \cos^2(\omega)C$$

$$= \sin^2(\omega)A + \sin(\omega)\cos(\omega)B + \cos^2(\omega)C.$$

The procedure of finding optimal perturbation is done by using the eigenvalues and eigenvectors of the above matrix function $P(\omega)$. From this the minimum and maximum eigenvalues of $P(\omega)$ are identified, which represent the critical frequencies $\hat{\omega}$. The corresponding eigenvector $\hat{x}$ is then used to construct the optimal perturbation $\Delta Q(\lambda)$.

If $Q(\lambda)$ is hyperbolic we need to find $\Delta Q(\lambda) = \lambda^2\Delta A + \lambda\Delta B + \Delta C$, with the smallest norm

$$\begin{pmatrix} \Delta A & \frac{\Delta B}{2} \\ \frac{\Delta B}{2} & \Delta C \end{pmatrix}$$

such that $Q(\lambda) + \Delta Q(\lambda)$ is not hyperbolic.

Theorem 43 ([15], Thm. 11)**.** *Let $Q(\lambda)$ be hyperbolic, then any perturbation $\Delta Q(\lambda)$ such that $Q(\lambda) + \Delta Q(\lambda)$ is not hyperbolic satisfies the inequality*

$$-r_H \leq \left\| \begin{pmatrix} \Delta A & \frac{\Delta B}{2} \\ \frac{\Delta B}{2} & \Delta C \end{pmatrix} \right\|_2 \leq \left\| \begin{pmatrix} \Delta A & \frac{\Delta B}{2} \\ \frac{\Delta B}{2} & \Delta C \end{pmatrix} \right\|_F,$$

where $r_H = \min_\omega \lambda_{max} P(\omega) < 0$. Furthermore, equality is achieved for the rank-one perturbations

$$\begin{pmatrix} \Delta A & \frac{\Delta B}{2} \\ \frac{\Delta B}{2} & \Delta C \end{pmatrix} = -r_H \left(\begin{bmatrix} \sin(\hat{\omega}) \\ \cos(\hat{\omega}) \end{bmatrix} \begin{bmatrix} \sin(\hat{\omega}) & \cos(\hat{\omega}) \end{bmatrix} \right) \otimes (\hat{x}\hat{x^*})$$

with $\hat{\omega} = \arg\min_\omega \lambda_{max} P(\omega)$ and $P(\hat{\omega})\hat{x} = r_H\hat{x}$ ($\| \hat{x} \|_2 = 1$).

On the other hand, if $Q(\lambda)$ is not hyperbolic we need to find $\Delta Q(\lambda) = \lambda^2\Delta A + \lambda\Delta B + \Delta C$, with the smallest norm

$$\begin{pmatrix} \Delta A & \frac{\Delta B}{2} \\ \frac{\Delta B}{2} & \Delta C \end{pmatrix}$$

such that $Q(\lambda) + \Delta Q(\lambda)$ is hyperbolic.

6.3.2 Overdamped Hyperbolic Quadratic Eigenvalue Problem

An important subclass of hyperbolic quadratic eigenvalue problem is overdamped hyperbolic QEP that arise in vibration problems of overdamped systems. The eigenvalues of an overdamped hyperbolic QEP can be efficiently computed exploiting the Hermitian and definiteness properties which guarantee real eigenvalues. The eigenvalues of overdamped hyperbolic QEP are necessarily real and nonpositive.

As the following equations show, the eigenvalues of overdamped $HQEP$ can be considered as shifted hyperbolic quadratic eigenvalues.

$$\begin{aligned}
Q(\lambda) &= \lambda^2 A + \lambda B + C \\
Q(\lambda+\theta) &= (\lambda+\theta)^2 A + (\lambda+\theta)B + C \\
&= (\lambda^2 + 2\lambda\theta + \theta^2)A + \lambda B + \theta B + C \\
&= \lambda^2 A + 2\lambda\theta A + \theta^2 A + \lambda B + \theta B + C \\
&= \lambda^2 A + \lambda(2\theta A + B) + C + \theta B + \theta^2 A \\
\tilde{Q}(\lambda) &= \lambda^2\tilde{A} + \lambda\tilde{B} + \tilde{C}
\end{aligned}$$

where $\tilde{A} = A > 0$, $\tilde{B} = B + 2\theta A > 0$ and $\tilde{C} = C + \theta B + \theta^2 A \geq 0$. This allows us to transform any hyperbolic QEP into an overdamped hyperbolic QEP.

Theorem 44. *A hyperbolic QEP is overdamped if and only if* $\lambda_1 \leq 0$.

Proof: See Guo and Lancaster [13]. □

7. CONCLUSION AND DISCUSSION

The quadratic eigenvalue problem has a staggering number of applications, for recent survey of the applications of (QEP) see [42]. The study investigates methods for solving generalized and quadratic eigenvalue problem with Hermitian matrices. Special emphasis is given to positive definite (GEP) problem and hyperbolic $(HQEP)$. The solution of QEP is a lot more difficult compared to linear eigenvalue problems such as generalized eigenvalue problem (GEP) and standard eigenvalue (SEP). However, the techniques used for solving linear eigenvalue problems can be used (with some modification) for solving the QEP. The linearization Methods transform $2n$ QEP to $2n \times 2n$ GEP. The linearization is not unique, therefore, a stable linearization is suggested. The Schur complement argument can be used to show that the GEP is a linearization of the QEP when A is invertible.

The $(HQEP)$ present nice properties such as having real eigenvalues and the coefficient matrices are Hermitian (or real symmetric). One of the methods used for testing for hyperbolicity is to first convert the QEP into GEP using one of the linearization techniques and then test the resultant Hermitian generalized eigenvalue problem $(HGEP)$ for positive definiteness using Crawford number. Once confirmed the hyperbolicity the eigenvalues of $HQEP$ are computed. As presented by [28], the $HQEP$ is transformed into positive definite generalized eigenvalue problem with twice the size of the original QEP.

The linearized GEP may be solved using the methods such as QZ method, Cholesky factorization or other iterative algorithms. For example if Y is nonsingular it is possible to calculate the few minimum and the few maximum eigenvalues using Lanczos method and Arnoldi method.

In the circumstances when the matrix pair of the GEP are Hermitian positive definite the GEP can be reduced to a standard eigenvalue problem (SEP) by conducting a structure preserving decomposition, namely simultaneous diagonalization by congruence. If the matrix pencil is not positive definite, a linear combination can be found such that $\alpha A + \beta B > 0$ is positive definite. The congruence transformation does not affect the nice property of Hermitian (or real symmetric). Once the GEP is transformed into SEP we can employ one of the iterative methods to compute all the eigenvalues (or extremal eigenvalues) as required [14].

BIBLIOGRAPHY

[1] D. Afolobi. Linearization of the quadratic eigenvalue problem. *Computers & Structures*, 26(6):1039–1040, 1987.

[2] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst. *Templates for the Solution of Algebraic Eigenvalue Problems, A Practical Guide.* SIAM, Philadelphia, 2000.

[3] C. S. Ballantine. Numerical range of a matrix: some effective criteria. *Linear Algebra and its Applications*, 19(2):117–188, 1978.

[4] L. Barkwell and P. Lancaster. Overdamped and gyroscopic vibrating systems. *Trans. AME: J. Applied Mechanics*, 59:176181, 1992.

[5] B. Bilir and C. Chicone. A generalization of the inertia theorem for quadratic matrix polynomials. *Linear Algebra and its Applications*, 280:229–240, 1998.

[6] D. A. Bini, L. Gemignani, and B. Meini. Computations with infinite Toeplitz matrices and polynomials. *Linear Algebra and its Applications*, 2002.

[7] D. A. Bini and V. Pan. Polynomial division and its computational complexity. *Journal of Complexity*, 2(3):179–203, 1986.

[8] D. A. Bini and V. Pan. *Matrix and Polynomial Computations*, volume 1 of *Fundamental Algorithms.* Birkhäuser, Boston, 1994.

[9] M.-D. Choi and C.-K. Li. Numerical range of the powers of an operator. *Journal of Mathematical Analysis and Application*, 365:458–466, 2009.

[10] P. I. Davies, N. J. Higham, and F. Tisseur. Analysis of the Cholesky method with iterative refinement for solving the symmetric definite generalized eigenproblem. *SIAM J. Matrix Anal. Appl.*, 23(2):472–493, 2001.

[11] W. Givens. Fields of values of a matrix. *Proc. American Mathematical Society*, 3:206–209, 1952.

[12] G. H. Golub and Ch. F. Van Loan. *Matrix computation.* The Johns Hopkins University Press, 1989.

[13] C. H. Guo and P. Lancaster. Algorithms for hyperbolic quadratic eigenvalue problems. *Mathematics of Computation*, 74:1777–1791, 2005.

[14] J.-S. Guo, W.-W. Lin, and Ch.-S. Wang. Numerical solution for large sparse quadratic eigenvalue problems. *Linear Algebra and its Applications*, 225:57–89, 1995.

[15] Y. Hachez and P. Van Dooren. Elliptic and hyperbolic quadratic eigenvalue problem and associated distance problems. *Linear Algebra and its Application*, 371:31–44, 2003.

[16] N. J. Higham, P. M. Van Dooren, and F. Tisseur. Detecting a definite Hermitian pair and a hyperbolic or elliptic quadratic eigenvalue problem, and associated nearness problems. *Linear Algebra and its Applications*, 351-352:455–474, 2002.

[17] N. J. Higham and H.-M. Kim. Solving a quadratic matrix equation by Newton's method with exact line searches. *SIAM J. Matrix Anal. Appl.*, 23(2):303–316, 2001.

[18] N. J. Higham, D. S. Mackey, and F. Tisseur. The conditioning of linearizations of matrix polynomials. *SIAM Journal of Matrix Analysis and Application*, 28(4):1005–1028, 2006.

[19] N. J. Higham, D. S. Mackey, and F. Tisseur. Scaling, sensitivity and stability in the numerical solution of quadratic eigenvalue problem. *International Journal for Numerical Methods in Engineering*, 73:344–360, 2008.

[20] R. A. Horn and Ch. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[21] R. A. Horn and Ch. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.

[22] Ch. R. Johnson. A Gersgorin inclusion set for the field of values of a finite matrix. *Proc. American Mathematical Society*, 41:57–60, 1973.

[23] Ch. R. Johnson. Gersgorin sets and the field of values. *Journal of Mathematical Analysis and Application*, 45(2):416–419, 1974.

[24] Ch. R. Johnson. Functional characterizations of the field of values and the convex hull of the spectrum. *Proceedings of the American Mathematical Society*, 61(2):201–204, 1976.

[25] D. Kincaid and W. Cheney. *Numerical Analysis.* Brooks Cole Publishing Company, 1991.

[26] M. G. Krein and M. A. Naimark. The method of symmetric and Hermitian forms in the theory of the separation of the roots of algebraic equations,. *Linear and Multilinear Algebra*, 10:265–308, 1981.

[27] C.-K. Li and R. Mathias. Generalized eigenvalues of a definite Hermitian matrix pair. *Linear Algebra and its Applications*, 271:309–321, 1998.

[28] Y. Lin and L. Bao. Block second order Krylov subspace methods for large scale quadratic eigenvalue problems. *Applied Mathematics and Computation*, 181:413–422, 2006.

[29] D. S. Mackey, N. Mackey, Ch. Mehl, and V. Mehrmann. Vector spaces of linearizations for matrix polynomials. *SIAM Jounal of Matrix Analysis and Application*, 28(4):971–1004, 2006.

[30] K. N. Majindar. Linear combinations of Hermitian and real symmetric matrices. *Linear Algebra and its Applications*, 25:95–105, 1979.

[31] A. S. Marcus. Introduction to spectral theory of polynomial operator pencils. *American Mathematical Society*, 171, 1988.

[32] M. Marcus and H. Minc. *A survey of matrix theory and matrix inequalities.* Dover Publ., New York, 1964.

[33] M. Marcus and N. Moyls. Field convexity of a square matrix. *Proc. of American Mathematical Society*, 6:981–983, 1955.

[34] R. Mathias. Perturbation Bounds for the Polar Decomposition. Unpublished manuscript, 1997.

[35] R. Mathias and C.-K. Li. A perturbation bound for definite pencils. *Linear Algebra Appl.*, 179:191–202, 1993.

[36] A. I. Mees and D. P. Atherton. Domains containing the field of values of a matrix. *Linear Algebra and its Application*, 26:289–296, August 1979.

[37] W. V. Parker. Characteristic roots and field of values of a matrix. *Proc. American Mathematical Society*, 57:103–108, 1951.

[38] N. J. Pullman. *Matrix Theory and its Application.* Marcel Dekker Inc., New York, 1976.

[39] H. Schneider and G. Ph. Barker. *Matrix and Linear Algebra.* Dover Publication, 1989.

[40] G. W. Stewart. Perturbation bounds for definite generalized eigenvalue problem. *Linear Algebra Appl.*, 23:69–83, 1973.

[41] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory.* Academic Press Limited, 1990.

[42] F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Review*, 43(2):235–286, 2001.

[43] J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, 1965.

[44] A. Wragg and C. Underhill. Remarks on the zeros of Bessel polynomials. *American Mathematical Monthly*, 83(2):122–126, 1976.

[45] Q. Ye. An iterated shift and invert arnoldi algorithm for quadratic matrix eigenvalue problem. *Applied Mathematics and Computation*, 172:818–827, 2006.