Matrix perturbations: bounding and computing eigenvalues

Reis da Silva, R.J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 1

Introduction

Perturbation, a disturbance of motion, course, arrangement, or state of equilibrium.
- Merriam-Webster dictionary

Despite the somewhat negative connotation of the word, not every perturbation is a bad perturbation. In fact, while disturbing the entries of a matrix, many perturbations still preserve useful properties such as the orthonormality of the basis of eigenvectors or the Hermicity of the original matrix.

In the coming chapters we will encounter several property-preserving perturbations. Some of these are as ordinary as Hermitian perturbations of Hermitian matrices. Some others, such as normality preserving perturbations, or those yielding approximations from below are less studied. Nevertheless, also these turn out to hide some charming results. For instance, the reader might be surprised to discover that the knowledge of a single eigenpair of a normal matrix is all that is needed to augment it with a row and a column while retaining its normality. Or, that among all the linear combinations of normal matrices, $tA + (1 - t)B$ with $t \in \mathbb{R}$, there are either two or infinitely many normal matrices.
1.1. PERTURBATION THEORY

We shall also encounter some pleasant perturbations. Those that affect entries in such a way that the properties of the resulting matrix are more gracious than the ones we started from. We find them, for instance, in the context of iterative methods for eigenvalue problems. These are perturbations that make the computation of the eigenvalues more efficient by introducing extra sparsity and/or a better structure. Determining general suitable perturbations for that purpose is still a difficult task. Nevertheless, in the framework of the Subspace Projected Approximate Matrix method some property-preserving perturbations turn out to be both suitable and to simplify the theoretical analysis.

Broadly speaking, however, this thesis is less about the perturbations themselves and more about the eigenvalues of the resulting matrix. Therefore, among the main results, the reader will find improved bounds for the eigenvalues of Hermitian perturbations of Hermitian matrices and new estimates for the eigenvalues of certain normality preserving perturbations of normal matrices.

1.1 Perturbation theory

Research on eigenvalue problems is one of the most prolific topics in Numerical Linear Algebra [31]. Perturbation analysis, in particular, studies the consequences of small changes in a matrix $A \in \mathbb{C}^{n \times n}$ to its eigenvalues and/or eigenvectors. The perturbation may result from the inherent inexact nature of numerical computations but also from research questions in the context of discretized differential equations, or stability of dynamical systems, to name a few. Independently of their origin, and assuming that the eigendata of $A$ are known, the problem is to determine the eigenvalues (and sometimes also eigenvectors) of the matrix

$$A + E$$

where $E \in \mathbb{C}^{n \times n}$ is some perturbation matrix. In many cases, however, rather than the exact eigenvalues of $A + E$, lower or upper bounds suffice and the deeper our knowledge of the characteristics of $A$ and $E$, the better can it be exploited to improve those bounds.
We may collect the results in perturbation theory for eigenvalues into two major groups. The first includes bounds for the error (absolute or relative) of the perturbation and examples of important results are Hoffman-Wielandt’s ([38], [39, Thm. 6.3.5]) and Bauer-Fike’s theorems. Because we shall refer to Bauer-Fike’s theorem often in the next chapters, we chose to present here also its proof.

**Theorem 1.1.1** (Bauer-Fike [5]). For \( n \in \mathbb{N} \), let \( \tilde{A}, A \) and \( E \) be \( n \times n \) matrices satisfying \( \tilde{A} = A + E \). Assume, furthermore, that \( V \in \mathbb{C}^{n \times n} \) is nonsingular and such that \( \Lambda = V^{-1}AV \) is diagonal. Finally, for \( j \in \{1, \ldots, n\} \) denote by \( \lambda_j \) the eigenvalues of \( A \) and let \( \mu \) be an eigenvalue of \( \tilde{A} \). Then, for at least one eigenvalue, \( \lambda_j \), of \( A \)

\[
|\lambda_j - \mu| \leq \kappa_p(V)\|E\|_p,
\]

where \( \kappa_p(V) = \|V^{-1}\|_p\|V\|_p \) is the condition number of \( V \) with respect to the matrix \( p \)-norm.

**Proof.** For notation simplicity, we define in this proof \( \|\cdot\| := \|\cdot\|_p \) for the matrix \( p \)-norm. If \( \mu \) is an eigenvalue of \( A \), it is enough to take \( \lambda_j = \mu \). We assume, therefore, that \( \mu \) is not an eigenvalue of \( A \). Owing to the fact that \( \mu \in \sigma(\tilde{A}) \), the matrix \( \tilde{A} - \mu I \) is singular and since \( \det(\Lambda - \mu I) \neq 0 \) we may write

\[
0 = \det(\tilde{A} - \mu I) = \det(\Lambda + V^{-1}EV - \mu I)
= \det(\Lambda - \mu I)\det \left[(\Lambda - \mu I)^{-1}V^{-1}EV + I\right].
\]

Thus, \( \det \left[(\Lambda - \mu I)^{-1}V^{-1}EV - (-1)I\right] = 0 \). This points to \(-1\) as an eigenvalue of \((\Lambda - \mu I)^{-1}V^{-1}EV \). Therefore,

\[
1 \leq \|(\Lambda - \mu I)^{-1}V^{-1}EV\| \leq \|V^{-1}\|_p\|E\|_p\|V\|
= \kappa_p(V)\|\Lambda - \mu I\|^{-1}\|E\|_p.
\]

Here, \( \kappa_p(V) = \|V^{-1}\|_p\|V\| \) and the first inequality follows from the fact that, for consistent norms, \( \|A\| \) is never smaller than the spectral radius of \( A \). This gives

\[
1 \leq \max |(\lambda_j - \mu)^{-1}|\|E\|\kappa_p(V) \quad \text{or equivalently}
min |\lambda_j - \mu| \leq \kappa_p(V)\|E\|_p.
\]
Then, for at least one \( j \), we have \( |\lambda_j - \mu| \leq \kappa_p(V)\|E\| \), proving the statement.

The other group of perturbation results focuses on the relative position of the eigenvalues of \( A \) and \( A + E \). Important results include Weyl’s inequalities on Hermitian perturbations of Hermitian matrices but also, in some sense, Cauchy’s Interlace Theorem. The two types of results are not mutually exclusive and they often appear together. Our interest, however, lies with this second group which we now develop in more detail.

1.1.1 Classical theory

Determining all values \( \lambda \) for which \( Av = \lambda v \) has a nonzero solution \( v \) is called the standard algebraic eigenvalue problem. Given such an eigenvector \( v \) of unit norm, we may use the same equation to determine the associated eigenvalue by writing \( \lambda = v^*Av \). For a matrix \( A \) (usually Hermitian), the quadratic form that relates an arbitrary vector in \( \mathbb{C}^n \) with a complex scalar is known as the Rayleigh quotient or Rayleigh-Ritz ratio.

**Definition 1.1.2** (Rayleigh quotient). For a matrix \( A \in \mathbb{C}^{n \times n} \) the Rayleigh quotient is the map

\[
\rho_A : \mathbb{C}^n \setminus \{0\} \rightarrow \mathbb{C} : x \mapsto \frac{x^*Ax}{x^*x}.
\]

For normal matrices, the Rayleigh quotient possesses several agreeable properties that include homogeneity, boundedness and translation invariance (see [62] or [63, §1.5]) and of which we emphasize one:

**Lemma 1.1.3** (Minimal Residual [62]). Given \( x \in \mathbb{C}^n \setminus \{0\} \), the scalar \( \rho_A(x) \) is the quantity that minimizes \( \|Ax - \mu x\|_2 \). Otherwise stated,

\[
\|Ax - \rho_A(x)x\|_2 \leq \|Ax - \mu x\|_2 \quad \text{for all} \quad \mu \in \mathbb{C}.
\]

**Proof.** Set \( \rho := \rho_A(x) \). Then,

\[
\|Ax - \mu x\|_2^2 = x^*x|\mu|^2 - \bar{\mu}x^*Ax - \mu x^*A^*x + x^*A^*Ax
\]

\[
= x^*x\left( x^*A^*Ax/x^*x + (\mu - \rho)(\bar{\mu} - \bar{\rho}) - |\rho|^2 \right)
\]

\[
\geq \|Ax\|_2^2 - |\rho|^2\|x\|_2^2.
\]
The statement is proved noting that \( \|Ax - \rho x\|^2_2 = \|Ax\|^2_2 - |\rho|^2 \|x\|^2_2 \).

Such properties make Rayleigh’s quotient an important tool not only in perturbation theory but also for the actual computation of eigenvalues. In fact, assuming that \( \|x\|_2 = 1 \), the norm of the spectral residual, \( \|Ax - \rho A(x)x\|_2 \), is a measure of how close a given pair, \((x, \rho A(x))\), formed by a vector and its Rayleigh quotient is to an exact eigenpair of \(A\). It is, therefore, often used as stopping criterion for iterative methods.

For a matrix \(A\), the subset of the complex plane resulting from considering the Rayleigh quotient of every nonzero vector \(x \in \mathbb{C}^n\) constrained to the unit \((n-1)\)-sphere is called the field of values of the matrix \(A\).

**Definition 1.1.4 (Field of values).** Given \(A \in \mathbb{C}^{n \times n}\), the field of values of \(A\) is the set
\[
F(A) = \{x^*Ax : \|x\|_2 = 1, x \in \mathbb{C}^n\}.
\]

The set \(F(A)\) is the image of the unit \((n-1)\)-sphere under the quadratic form \(x^*Ax\) and is closed, bounded and convex in \(\mathbb{C}\). During the last sixty years, several generalizations of the field of values were proposed [40, §1.8]. Examples include Givens field of values generalizing the inner-product [29]; Bauer field of values with different right and left vectors [4, 58, 85]; the \(q\)-field of values [53, 79]; and more recently the generalized two-sided field of values [66] by the author of this thesis. Each generalization attempts to replicate one or more properties of the standard field of values.

If \(A\) is a normal matrix, \(F(A)\) is the convex hull of the eigenvalues of \(A\). In particular, if \(A\) is Hermitian, the field of values is a real interval. In this case, we are able to express the largest and smallest eigenvalues of \(A\) as solutions of constrained optimization problems involving \(F(A)\).

**Theorem 1.1.5 (Rayleigh-Ritz [39]).** If \(A \in \mathbb{C}^{n \times n}\) is a Hermitian matrix with eigenvalues \(\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n\) then
\[
\lambda_1 = \min_{x \in \mathbb{C}^n \setminus \{0\}} \frac{x^*Ax}{x^*x} = \min_{x \in \mathbb{C}^n, \|x\|_2 = 1} x^*Ax,
\]
\[ \lambda_n = \max_{x \in \mathbb{C}^n \setminus \{0\}} \frac{x^*Ax}{x^*x} = \max_{x \in \mathbb{C}^n, \|x\|_2 = 1} x^*Ax \]

and consequently,

\[ \lambda_1 \|x\|_2^2 \leq x^*Ax \leq \lambda_n \|x\|_2^2, \quad \forall x \in \mathbb{C}^n. \]

**Proof.** Let \( n \in \mathbb{N} \) and let \( V \in \mathbb{C}^{n \times n} \) be unitary and such that \( \Lambda = V^*AV \) is diagonal. Recall, that for any vector \( x \in \mathbb{C}^n \) we have \( \|V^*x\|_2^2 = x^*VV^*x = x^*Ix = \|x\|_2^2 \). Consequently,

\[ x^*Ax = (V^*x)^*\Lambda(V^*x) = \sum_{j=1}^{n} \lambda_j |(V^*x)_j|^2. \]

Now, replacing every \( \lambda_j \) by either \( \lambda_{\text{min}} \) or \( \lambda_{\text{max}} \) we obtain the inequalities

\[ \lambda_{\text{min}} \sum_{j=1}^{n} |(V^*x)_j|^2 \leq \lambda_j \sum_{j=1}^{n} |(V^*x)_j|^2 \leq \lambda_{\text{max}} \sum_{j=1}^{n} |(V^*x)_j|^2 \]

or differently stated

\[ \lambda_1 \|x\|_2^2 \leq x^*Ax \leq \lambda_n \|x\|_2^2. \]

For the remaining statements, note that if \( Ax = \lambda_j x \) then \( x^*Ax = \lambda_j \|x\|_2^2 \). \( \square \)

If \( A \) is not normal, however, \( F(A) \) is, generally, no longer the convex hull of \( \sigma(A) \) and consequently it is not maximized or minimized in the eigenvectors associated with the largest and smallest eigenvalues of \( A \), even if these are real.

In fact, not only the largest and smallest but every eigenvalue of a Hermitian matrix \( A \) can be expressed in terms of a constrained optimization problem. This result is known as the Courant-Fischer minimax theorem \([20, 25]\) whose proof we present here following the approach from Parlett \([63\text{ Thm. 10.2.1}]\).

**Theorem 1.1.6 (Courant-Fischer [63]).** Let \( A \in \mathbb{C}^{n \times n} \) be Hermitian and let \( W^k \) denote a subspace of \( \mathbb{C}^n \) with dimension indicated by the superscript. Then, the \( j \)th eigenvalue of \( A \) in non-decreasing order of magnitude satisfies

\[ \lambda_j(A) = \min_{W^j} \max_{0 \neq x \in W^j} \rho_A(x) = \max_{W^{n-j+1}} \min_{0 \neq u \in W^{n-j+1}} \rho_A(u) \]
Proof. The dimensions of $W^j$ and $W^{n-j+1}$ and the basis theorem for finite dimensional spaces ensure that $W^j \cap W^{n-j+1} \neq \{\vec{0}\}$. Denote by $z$ any nonzero vector in $W^j \cap W^{n-j+1}$. Then,

$$\min_{\vec{0} \neq u \in W^{n-j+1}} \rho_A(u) \leq \rho_A(z,z) \leq \max_{\vec{0} \neq x \in W^j} \rho_A(x).$$

Because the inequalities are valid for all choices of $W^j$ and $W^{n-j+1}$, we may take the maximum over all of them on the left-hand side and the minimum on the right-hand side leading to

$$\max_{W^{n-j+1}} \min_{\vec{0} \neq u \in W^{n-j+1}} \rho_A(u) \leq \min_{W^j} \max_{\vec{0} \neq x \in W^j} \rho_A(x).$$

Recall that, if $x = u = v_k$ is an eigenvector of $A$ associated with eigenvalue $\lambda_k$, its Rayleigh quotient satisfies $\rho_A(v_k) = \lambda_k$. Therefore, equality follows from choosing $W^j = \text{span}\{v_1, \ldots, v_j\}$ and $W^{n-j+1} = \text{span}\{v_j, \ldots, v_n\}$ where for each $k \in \{1, \ldots, n\}$ each $v_k$ is an eigenvector of $A$ associated with eigenvalue $\lambda_k$.

### 1.1.2 Perturbation results

Because when $A$ and $E$ are Hermitian so is $A + E$, Courant-Fischer’s theorem may be used to estimate its eigenvalues in terms of those of $A$ and $E$. This is the statement of the next theorem, originally presented by Weyl.

**Theorem 1.1.7** (Weyl’s inequalities [81]). Let $A$ and $E$ be $n \times n$ Hermitian matrices. For $j \in \{1, \ldots, n\}$ denote by $\lambda_j, \mu_j$ and $\delta_j$ the $j$th eigenvalue of $A + E$, $A$ and $E$ respectively. Assume these to be in non-decreasing order of magnitude. Then,

$$\mu_j + \delta_1 \leq \lambda_j \leq \mu_j + \delta_n.$$

**Proof.** For any vector $x \in \mathbb{C}^n$, Theorem 1.1.5 guarantees that

$$\delta_1 x^* x \leq x^* E x \leq \delta_n x^* x.$$

Therefore, using Courant-Fischer’s theorem (Theorem 1.1.6), we have for each
$j \in \{1, \ldots, n\}$

\[
\lambda_j = \min_{W_j \subseteq \mathbb{C}^n} \max_{\tilde{\delta} \neq x \in W_j} \frac{x^* (A + E) x}{x^* x} = \min_{W_j \subseteq \mathbb{C}^n} \max_{\tilde{\delta} \neq x \in W_j} \left( \frac{x^* Ax}{x^* x} + \frac{x^* Ex}{x^* x} \right) \\
\geq \min_{W_j \subseteq \mathbb{C}^n} \max_{\tilde{\delta} \neq x \in W_j} \left( \frac{x^* Ax}{x^* x} + \delta_1 \right) = \mu_j + \delta_1.
\]

The remaining inequality is obtained using similar arguments.

Theorem 1.1.7 provides us with bounds for the eigenvalues of $A + E$ in terms of the eigenvalues of $A$ and $E$. Because the eigenvalues of $A$ are assumed to be known, the usefulness of Theorem 1.1.7 depends on how much is known of $E$. For instance, if $E$ is a rank-one Hermitian positive semidefinite matrix, Theorem 1.1.7 simplifies into the following Corollary.

**Corollary 1.1.8.** Let $E = ww^*$ be a Hermitian rank-one positive semidefinite matrix. Then,

\[
\mu_j \leq \lambda_j \leq \mu_j + \|w\|_2^2.
\]

Note that Corollary 1.1.8 may, equivalently, be stated in terms of a rank-one negative semidefinite matrix $E$. In either situation, if the eigenvalues of $A$ are assumed to be known, this results in easily computable bounds for the eigenvalues of $A + E$. Corollary 1.1.8 is of interest on its own, but also because perturbing a Hermitian matrix $A$ with a Hermitian matrix $E$ can, at least in theory, be studied from a succession of perturbations of $A$ by Hermitian rank-one matrices.

**Recent contributions**

The following more recent contributions on perturbation theory will prove useful for the contextualization of the results from Chapter 2.

A renewed interest in the analysis of rank-one perturbations of diagonal matrices grew with the development of the divide-and-conquer algorithm for the computation of the eigenvalues of Hermitian tridiagonal matrices. The algorithm divides a tridiagonal matrix into a sum of a block diagonal matrix with tridiagonal blocks and a Hermitian rank-one matrix. Given the diagonalization of the submatrices, the remaining problem is that of determining the eigenvalues of a
Hermitian rank-one perturbation $E = \omega \sigma \omega^*$ of a diagonal matrix $D$. If the eigenvalues, $\delta_1, \ldots, \delta_n$, of $D$ are distinct and $w_j \neq 0$, then solving $(D + E)v = \lambda v$ with $w^*v \neq 0$ is equivalent to solving $1 + \sigma w^*(D - \lambda I)^{-1}w = 0$ or, in scalar form, to determining the roots of the so-called secular equation:

$$f(\lambda) = 1 + \sigma \sum_{j=1}^{n} \frac{w_j^2}{\delta_j - \lambda}. \quad (1.1)$$

The eigenvalues of a diagonal plus a rank-one matrix are usually computed using methods based on the secular equation [15, 51, 54, 59, 83]. In Section 5 of [30], Golub describes alternative methods for the eigenvalue problem of a rank-one perturbation of a diagonal matrix. His approach is to transform the original problem into one for which efficient algorithms already exist. Methods based on the secular equation (among others) are mentioned and an alternative $O(n^2)$ algorithm is devised. Bunch et al. [15] analyze the computation of the eigenvalues of rank-one perturbations of general Hermitian matrices. Golub’s results (from [30]) are extended to include the eigenvector computation as well as the possibility of the matrix having multiple eigenvalues.

Weyl’s inequalities, used to bound the eigenvalues of $A + E$ from below by the eigenvalues of $A$, are among the best known and easily computable results. More recently, however, Ipsen and Nadler [46] show how Weyl’s bounds can be improved at a small cost. These authors bound the smallest eigenvalues of $A + E$ from below by the smallest eigenvalue of a $2 \times 2$ eigenproblem resulting from clever manipulations of the matrices $A$ and $E$. Making use of Cauchy’s Interlace Theorem (see Theorem 1.1.10) they propose, also, lower bounds for the interior eigenvalues of $A$. Additional bounds are given for (well-separated) eigenvalues of Hermitian matrices when the perturbation is Hermitian or non-Hermitian with sufficiently small norm.

Perturbation problems arise not only for the eigenvalue but also for the singular value problem. Given a matrix $B$ with known singular values, one wishes to determine the singular values of some perturbed matrix $\tilde{B}$. A survey in perturbation theory for the Singular Value Decomposition is given by Stewart in [76, 77]. It covers results ranging from singular value bounds to low rank approximations
and singular vectors perturbation. Applications in the field of image analysis and information retrieval can be found in [16] and [78] respectively. Another frequent problem occurs when $\tilde{B}$ is the matrix $B$ after columns or rows were added or removed. Bunch and Nielsen [14] studied this situation and developed algorithms for the efficient computation of the updated singular value decomposition as well as for the updated least squares problem.

1.1.3 Normal matrices

As soon as we leave the safe haven of Hermitian matrices, perturbation results and, in particular, perturbation results in terms of the relative position of the eigenvalues, become scarce. One reason is that for general normal matrices, those whose basis of eigenvectors is still orthonormal, many fundamental questions are still unanswered. We contextualize the results from Chapter 3 with some background on perturbation results for normal matrices.

The location of the eigenvalues of sums of normal matrices has been studied as early as 1955 by Wielandt [82]. He did not, however, required their sum to be normal. Another commonly studied subclass of normal matrices is that of essentially Hermitian matrices [7, 23]. Essentially Hermitian matrices are matrices $E$ that can be written as

$$E = \beta H + \alpha I,$$

where $\alpha, \beta \in \mathbb{C}$ and where $H$ is Hermitian.

Faber and Manteuffel show in [24] that for linear systems with essentially Hermitian system matrix, there exists a variant of the Conjugate Gradient method that still relies on a three term recursion. In the context of the Arnoldi method applied to normal matrices, Huckle [41] proves that irreducible normal tridiagonal matrices are essentially Hermitian. In fact, the upper Hessenberg matrix generated by the Arnoldi method is tridiagonal only for essentially Hermitian matrices. The Arnoldi’s method can, in that case, be simplified to the Lanczos method.

In [42], Huhtanen shows that for every normal matrix $A$ and almost all unimodular $z \in \mathbb{C}$, the skew-Hermitian part $S(zA)$ of $zA$ is a polynomial $p$ of degree at most $n - 1$ in the Hermitian part $H(zA)$. The essentially Hermitian matrices
are exactly the ones for which the degree of \( p \) is at most one, resulting in collinear eigenvalues. Since \( p \) can be retrieved in a modest number of arithmetic operations as a by-product of the Lanczos algorithm applied to \( \mathcal{H}(zA) \), this led to the development \cite{12, 43} of efficient structure preserving algorithms for eigenvalue problems and linear systems involving normal matrices.

In \cite{44}, Ikramov investigates which matrices \( R \) can be the upper triangular part of a normal matrix \( A \). Although the full problem remains open, a characterization exists if all diagonal entries of \( R \) lie on the same line in the complex plane, say

\[
\ell : \{ \gamma + \theta \rho \mid \rho \in \mathbb{R} \}, \quad \gamma, \theta \in \mathbb{C} \text{ with } |\theta| = 1.
\]

Writing \( R = D + U \) where \( D \) is the diagonal of \( R \), we see that \( \tilde{\theta}(D+U-\gamma I) \) is upper triangular with real diagonal and, consequently, that \( H = \tilde{\theta}(D+U-\gamma I) + \theta U^* \) is Hermitian. Therefore,

\[
A = \theta H + \gamma I = R + \theta^2 U^*
\]

is essentially Hermitian, hence normal, with upper triangular part equal to \( R \).

Ikramov and Elsner also studied the normal augmentation problem which is a particular case of the topic of the next section.

### 1.1.4 Augmentation of matrices

A problem related to perturbation of matrices is the one of augmenting a matrix with one or more rows and columns. Formally, we may state it as follows. *Given* \( A \in \mathbb{C}^{n \times n} \), *characterize all* \( n \times m \) *matrices* \( V, W \in \mathbb{C}^{n \times m} \), *and all* \( \Gamma \in \mathbb{C}^{m \times m} \) *such that*

\[
A_+ = \begin{bmatrix} A & V \\ W^* & \Gamma \end{bmatrix},
\]

*(1.2)*

*satisfies certain properties.*

The matrix \( A \) is called a *principal submatrix* of \( A_+ \).

**Definition 1.1.9** (Principal submatrix). *A* \( k \times k \) *matrix* \( A \) *is called a principal submatrix* of \( A_+ \) *if* \( A \) *is obtained from* \( A_+ \) *by removing* \( n - k \) *columns and their corresponding rows.*
Note that we may obtain principal submatrices by removing consecutive rows and columns from the top left, as well as from the bottom right corners of $A_+$. The resulting principal submatrices are called leading and trailing respectively.

If $A$ is Hermitian then $A_+$ is also Hermitian if and only if $W = V$ and $\Gamma = \Gamma^*$. In this case, Cauchy Interlace Theorem relates the eigenvalues of the Hermitian matrix $A_+$ with those of $A$.

**Theorem 1.1.10 (Cauchy Interlace [63]).** Let $A \in \mathbb{C}^{n \times n}$ be Hermitian and let $V$ and $\Gamma$ be such that

$$A_+ = \begin{bmatrix} A & V \\ V^* & \Gamma \end{bmatrix}$$

is Hermitian. Denote by

$$\mu_1 \leq \mu_2 \leq \ldots \leq \mu_n \quad \text{and} \quad \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{n+m}$$

the eigenvalues of $A$ and $A_+$ respectively. Then

$$\lambda_j \leq \mu_j \leq \lambda_{j+m} \quad \text{for all} \quad j \in \{1, \ldots, n\}.$$ 

In short, Cauchy’s Interlace Theorem states that the eigenvalues of $A$ are interlaced by those of any of its submatrices. The proof does not add relevant information. We therefore only present the result, and direct the reader to [63 §10.1] or [39 Thm 4.3.8] for two possible proofs.

If $A$ is normal and we want $A_+$ to be normal too the situation is more difficult. Ikramov and Elsner [45] dealt with this normality preserving augmentation for $m = 1$ and $m = 2$. In particular for $m = 1$, with the matrices $V, W$ and $\Gamma$ reduced to vectors $v$ and $w$ and a scalar $\gamma$, they established that for an arbitrary complex number $\xi$ of modulus one, $v$ must be an eigenvector of the matrix $A - \xi A^*$ while the vector $w$ equals $w = \xi v$. The scalar $\gamma$ is then dependent on the eigenvalue of $A - \xi A^*$ associated with $v$. The problem in its full generality, however, remains unsolved.

As a final remark on this subject, note that for the case that $m = 1$ and the matrix $A$ is diagonal and $v$ and $w$ are nonzero vectors, then the matrix $A_+$ is a so-
called 1-bordered diagonal matrix. The literature pertaining to bordered diagonal matrices is not extensive. Moreover, the literature that exists is ambiguous due to the use of the terms arrow matrix (for instance [2]) and arrowhead matrix (for instance [64]) to refer to the same type of matrices, namely those which, possibly after a permutation, are zero except for the diagonal and the last row and column. The most commonly used definition and the one with a more general scope seems to be $k$-bordered diagonal matrix as it includes the possibility of more than a single nonzero row and column. We shall, therefore, also adhere to this terminology.

Wilkinson [83] and Parlett [63] work with bordered diagonal matrices in the setting of perturbation theory for eigenvalues while O’Leary and Stewart [59] showed there are explicit expressions for the determinant and for the eigenvectors of symmetric 1-bordered diagonal matrices (arrowhead matrices). More recently, Parlett and Strang [64] studied arrowhead matrices with prescribed Ritz values.

### 1.2 Iterative methods for Eigenvalue Problems

Although every method to obtain the eigenvalues of a matrix is, in essence, iterative, the characterization as iterative is usually restricted to those that determine only a few eigenvalues (and eigenvectors) of large and usually sparse matrices [80]. These methods restrict the matrix $A$ to a suitable subspace and search for approximations to the eigenvalues of $A$ in that subspace. For this reason, they are also called subspace methods. Important concepts in this context are those of Ritz values and Ritz vectors.

#### 1.2.1 Ritz values and vectors

Let $V$ be an $n \times k$ matrix with mutually orthonormal columns and define a $k$-dimensional subspace of $\mathbb{C}^n$ as

$$\mathcal{V} = \{Vy \mid y \in \mathbb{C}^k\}.$$

In the context of iterative methods, we shall call the subspace $\mathcal{V}$ as the search subspace. Consider, in addition, a matrix $V_\perp$ whose columns form an orthonormal
basis for the orthogonal complement of \( V \). Then, \((V|V\perp)\) is unitary and an arbitrary Hermitian matrix \( A \) may be represented with respect to the basis given by the columns of \( V \) as

\[
\hat{A} = (V|V\perp)^* A (V|V\perp) = \begin{bmatrix} M & R^* \\ R & S \end{bmatrix}.
\]

(1.3)

In Equation (1.3), \( M = V^* AV \) is \( k \times k \), \( R \) is \((n - k) \times k \) and \( S = V^*_\perp A V^*_\perp \) is \((n - k) \times (n - k) \). Since \( A \) is Hermitian so is \( M \) and its \( k \) eigenvalues,

\[
\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{k-1} \leq \mu_k,
\]

(1.4)

are called the Ritz values of \( A \) with respect to \( V \). Any vector \( u = Vz \), where \( z \) is an eigenvector of \( M \) belonging to \( \mu_j \) is called a Ritz vector of \( A \) in \( V \), and \((\mu_j, u)\) a Ritz pair. Each Ritz pair can be interpreted as an eigenpair of the \( n \times n \) rank-\( k \) matrix \( VMV^* \) and is seen as an approximation of an eigenpair of \( A \). See \([47, 63, 71]\).

An additional remark is in order. If we were to add an extra column, \( v \), to \( V \), the new matrix \( M_{k+1} = V_{k+1}^* A V_{k+1} \) would be

\[
M_{k+1} = \begin{bmatrix} M & V^* Av \\ v^* AV & v^* Av \end{bmatrix}.
\]

(1.5)

The matrix \( M \) is a principal submatrix of \( M_{k+1} \). When \( A \) is Hermitian, both \( M \) and \( M_{k+1} \) are also Hermitian and from Cauchy Interlace Theorem (Theorem 1.1.10) we know that the eigenvalues of \( M \) interlace those of \( M_{k+1} \). Moreover, since \( A \) and \( \hat{A} \) have the same eigenvalues, the eigenvalues of \( M \) also interlace the ones of \( A \).

Related to Ritz values and vectors is another type of important approximate eigenpairs: harmonic Ritz values and vectors. In the context of iterative methods, and as the number of iterations increases, the Ritz values tend to approximate the eigenvalues located at the extrema of the spectrum of \( A \) better than those located in the interior \([63, \S 11]\). To approximate interior eigenvalues, Morgan \([56]\) suggested the use of Ritz pairs associated with the inverse of \( A \).
Definition 1.2.1 (Harmonic Ritz value). The harmonic Ritz values of $A$ with respect to the linear subspace $V$ are the reciprocals of the Ritz values of $A^{-1}$ with respect to $AV$.

Paige, Parlett and van der Vorst have (formally) introduced harmonic Ritz values for Hermitian matrices in [60]. In that publication, they make the connection between harmonic Ritz values and Lehmann’s optimal error bounds [50] for the eigenvalues of $A$ and show how, for Krylov subspaces, harmonic Ritz values can be cheaply obtained as by-products of the standard computation.

Other work on harmonic Ritz values and vectors include [73] from Sleijpen and van der Vorst introducing harmonic Ritz values for general matrices, [6] with an extensive study on their relation with Lehmann’s optimal bounds; [17] and [71] looking into a priori and a posteriori error analysis; [33] in the context of non-symmetric linear systems; and [35, 37] with generalizations to different eigenvalue problems.

1.2.2 Rayleigh-Ritz procedure

Traditionally, iterative methods are seen as working in two different stages (together forming one iteration) which are repeated until a criterion is met. The first includes the process of projecting the matrix in the search subspace and retrieving the Ritz pairs. At this stage, the matrix $S$ from (1.3) is not required. On the other hand, since the product $AV$ was already evaluated to compute $M$, a cheaply available by-product of the process is the matrix $\hat{R} = V_{\perp}R = AV - VM$. The columns $\hat{r}_j$ of $\hat{R}$ are the residuals for the respective Ritz pairs $(\mu_j, u_j)$ and satisfy

$$Au_j - \mu_ju_j = \hat{r}_j \perp V.$$

At each iteration, one or more Ritz pairs are chosen as approximations for the target eigenpair(s) of the original matrix. The norm of their residuals gives a measure of how close those pairs are to eigenpairs of $A$. Indeed, $(A - \hat{r}_ju_j^*)u_j = \mu_ju_j$ and thus, with $E = -\hat{r}_ju_j^*$, the Bauer-Fike Theorem (Theorem 1.1.1) shows that

$$|\mu_j - \lambda| \leq \|\hat{r}_j\|_2$$

(1.6)
for at least one eigenvalue $\lambda$ of $A$. Note, moreover, that line 11 in Algorithm 1 ensures that a sufficiently close approximation is found in the sense that $|\mu_j - \lambda| \leq \epsilon$ for some $\lambda \in \sigma(A)$ and some $\epsilon \in \mathbb{R}$.

If the Ritz values and vectors selected do not approximate the target eigenpair well enough, it is necessary to expand the search subspace. This leads to the second stage where an attempt is made to obtain a suitable vector with which to expand $V$. Different approaches for that selection distinguish the different iterative methods. This is indicated in line 14 of Algorithm 1 for general iterative methods for eigenvalue problems. In the coming sections we review three of these expansion strategies each leading to an iterative method for eigenproblems. These will be the Lanczos [49], Jacobi-Davidson [72] and Riccati [9] methods.

Algorithm 1 General subspace method

1: Input $A$, $v_0$, $\epsilon$;
2: for $q = 1, 2, \ldots$ do
3: \hspace{0.5cm} $v_q = v_{q-1}/\|v_{q-1}\|_2$;
4: \hspace{0.5cm} $w_q = Av_q$;
5: \hspace{0.5cm} for $p = 1, \ldots, q - 1$ do
6: \hspace{1cm} $M_{p,q} = v_p^*w_q$; $M_{q,p} = M_{p,q}^*$;
7: \hspace{0.75cm} end for
8: \hspace{0.5cm} $M_{q,q} = v_q^*w_q$;
9: Compute the target eigenpair $(\mu, z)$, with $\|z\|_2 = 1$, of $M$;
10: $u = Vz$; \{Ritz vector\}
11: $\hat{r} = Au - \mu u$; \{residual\}
12: if $\|\hat{r}\|_2 \leq \epsilon$ then
13: \hspace{0.5cm} stop; \{$(\mu, u)$ is the desired approximate eigenpair\}
14: else
15: \hspace{0.5cm} Obtain $t$; \{Specific to the different methods\}
16: \hspace{0.25cm} end if
17: Orthogonalize $t$ to the columns of $V$
18: end for

1.2.3 Lanczos method

Given a Hermitian matrix $A \in \mathbb{C}^{n \times n}$ and an initial search space $V_1 = \text{span}\{v_1\}$ with $\|v_1\|_2 = 1$, the Lanczos method [49] constructs an orthonormal basis for the
Krylov subspace

\[ \mathcal{V}_j = \mathcal{K}^j(A, v_1) = \text{span}\{v_1, Av_1, \ldots, A^{j-1}v_1\}. \]

If we orthogonalize the second basis vector \( Av_1 \) with respect to \( v_1 \), the result is \( Av_1 - (v_1^* Av_1) v_1 = \hat{r} \), i.e., the residual associated with the Ritz pair of the current iteration. Thus, assuming that \( \hat{r} \neq 0 \), \( \{v_1, \hat{r}\} \) is an orthogonal basis of \( \mathcal{V}_2 \). For higher dimensions, the expansion is made with the residual, \( \hat{r} \), associated with any of the Ritz pairs of \( A \) in \( \mathcal{V}_j \) and \( \mathcal{V}_{j+1} = \mathcal{V}_j \oplus \text{span}\{\hat{r}\} \). Since all residuals are multiples of one another, this uniquely defines the method.

In each step, a single matrix-vector product with the matrix \( A \) is required and each new expansion vector is normalized and appended to a matrix \( V_j \), whose columns span \( \mathcal{V}_j \). A consequence of the use of orthonormal bases for the Krylov subspaces is that \( M \) is tridiagonal. Formally, we may express this by the Lanczos relation

\[ AV_j = V_{j+1} M_{j+1,j}, \quad \text{with} \quad M_{j+1,j} = \begin{bmatrix} M_j \\ \tau e_j^* \end{bmatrix}. \]

Here, \( M_j \) is a \( j \times j \) tridiagonal matrix, \( e_j \) is the \( j \)th canonical basis vector of \( \mathbb{C}^j \) and \( \tau \) is a scalar.

A method mathematically equivalent to the Lanczos method can be implemented in terms of Algorithm 1 by replacing line 14 with \( t = \hat{r} \). However, due to the special structure of the subspaces it generates, it is possible to obtain a simpler implementation which we present in Algorithm 2.

The simplicity of the Lanczos method makes it a widely studied method. A selection of the most important references as well as a clear and relatively concise discussion of its mathematical properties, algorithmic implementations and variants is given in [3, 32]. The use of Lanczos methods for problems resulting from industrial applications is discussed in [48].

1.2.4 Jacobi-Davidson and Riccati methods

An alternative to Krylov subspaces was proposed by Davidson [22] in the method with his name. Rather than expanding the subspace \( \mathcal{V} \) with the residual, Davidson
Algorithm 2 Lanczos method

1: Input: $A, v_0, \epsilon$
2: $\beta_0 = \|v_0\|_2$
3: for $j = 1, 2, \ldots$ do
4: $v_j = v_{j-1}/\beta_{j-1}$
5: $w = Av_j$
6: $w = w - v_{j-1}\beta_{j-1}$
7: $\alpha_j = v_j^*w$
8: $v_j = w - v_j\alpha_j$
9: $\beta_j = \|v_j\|_2$
10: Compute the target eigenpairs of $M_j$; \{Ritz pairs\}
11: Compute the residual, $\hat{r}$
12: if $\|\hat{r}\|_2 \leq \epsilon$ then
13: stop
14: end if
15: end for

suggested to use the solution of the linear system

$$(D - \lambda I)t = -\hat{r}$$

where $D$ stands for the diagonal of the matrix $A$ to obtain a suitable expansion vector. This idea was improved in the Jacobi-Davidson method by Sleijpen and van der Vorst \[72\] and in the Riccati method from Brandts \[9\]. Just like Davidson’s method, Jacobi-Davidson and Riccati methods make use of subspaces without Krylov structure. Unlike in Davidson’s method, however, the new expansion vectors, are sought in the orthogonal complement of the current Ritz vector.

Given a Ritz pair $(\mu, u)$ with residual $\hat{r} \perp V$ and an $n \times (n-1)$ matrix $U_\perp$ such that $(u|U_\perp)$ is unitary, we may transform $A$ as

$$(u|U_\perp)^*A(u|U_\perp) = \begin{bmatrix} \mu & r^* \\ r & S \end{bmatrix},$$

where $r = U_\perp^*\hat{r}$ and $S = U_\perp^*AU_\perp$. Associated with the vector $u \in V$ there is, generally, a unique vector $t \in V_\perp$ such that $u + t$ is an eigenvector of $A$ corresponding to the target eigenvalue. For that reason and the fact that $t \perp u$, the vector $t$ is a so-called orthogonal correction to $u$. Moreover, $t = U_\perp y$ where $y$ can be verified.
to satisfy the generalized algebraic Riccati equation

\[(S - \mu I)y = -r + yr^*y, \quad (1.8)\]

or transformed back to the original basis

\[(I - uu^*)(A - \mu I)(I - uu^*)t = -\hat{r} + t\hat{r}^*t \quad \text{with} \quad t \perp u. \quad (1.9)\]

Solving Equation (1.9), however, is as difficult as solving the original eigenvalue problem. The idea in Jacobi-Davidson is, therefore, to first linearize Equation (1.9) and only then to solve it. The result is Jacobi-Davidson's correction equation

\[(I - uu^*)(A - \mu I)(I - uu^*)t = -\hat{r} \quad \text{with} \quad t \perp u \quad (1.10)\]

which has a unique solution and is, in practice, only solved approximately. The Riccati method, on the other hand, approximates the solutions of Equation (1.9) by means of a new Rayleigh-Ritz procedure in a smaller, \(\ell\)-dimensional subspace \(X\) of \(U^\perp\). A suitable solution is then chosen as expansion vector for \(V\). The Riccati method can, in this sense, be seen as an enhanced version of Jacobi-Davidson.

Both Jacobi-Davidson and Riccati methods can be implemented using Algorithm 1 with line (14) replaced by

**Solve (approximately)** \((I - uu^*)(A - \mu I)(I - uu^*)t = -\hat{r} \quad \text{for} \quad t \perp u;\)

in the Jacobi-Davidson method and with

**Solve (approximately)** \((I - uu^*)(A - \mu I)(I - uu^*)t = -\hat{r} + t\hat{r}^*t \quad \text{for} \quad t \perp u;\)

**Choose the minimum norm solution;**

in the Riccati method.

Both methods can be adapted to exploit extra available information such as, for instance, approximations for the matrix \(A - \mu I\). This may be seen as a form of preconditioning. This is only useful, however, if the approximation reduces the computational costs while preserving or even improving the rate of convergence towards the solution. Therefore, also for eigenvalue problems, determining good preconditioners is, in the words of Saad [68], “a combination of art and science”.

Recent contributions

A study of inner-outer iteration methods in the context of eigenvalues problems is given in [26]. Jacobi-Davidson and Riccati are compared with more detail in [10] and the analysis is complemented with some numerical experiments highlighting the differences between the two. The use of alternative correction equations for Jacobi-Davidson was studied in [28] and included, in particular, a projection onto the orthogonal complement of the entire current search subspace rather than only the current Ritz vector. Other variants of Jacobi-Davidson are suggested, for instance, in [36] for non-normal matrices and [70] for different types of eigenproblems.

To conclude this introduction, and to improve the overall readability of the remainder of this thesis, we now provide an outline of the different chapter contents and a few notational remarks.

Outline

Hermicity preserving perturbations are a standard choice in the study of eigenvalues bounds and we shall investigate them in Chapter 2. Starting from the study of a simple Hermitian rank-one perturbation $E$ of a Hermitian matrix $A$, we construct a sequence of increasingly better computable bounds for the eigenvalues of $A + E$. The advantage of this approach is that the accuracy of the bounds is dependent on the available eigendata of $A$ and, therefore, is potentially better than that of the classical solutions and the recent work in [46]. Moreover, the techniques suggested extend easily to other eigenvalue and singular value perturbation problems. The contents of this chapter, with the exception of Section 2.6, have been published as [11]:


Unlike the preservation of Hermicity, preserving normality is, in every sense, a complex problem. Chapter 3 makes headways on normality preserving pertur-
bations and augmentations of normal matrices. The latter situation has been addressed earlier by Ikramov and Elsner [45]. However, by recognizing the role played by the eigendata of the original matrix, we not only clarify their solution but also bound the eigenvalues of the augmented matrix. We also study normality preserving perturbations. For $2 \times 2$ and for rank-one matrices the analysis is now complete while for higher rank all essentially Hermitian normality perturbations are described. With this new information and using the stratification of normal matrices suggested by Huhtanen [42] in 2001 we construct all augmentations that result in normal matrices with eigenvalues on a quadratic curve in the complex plane. The contents of this chapter may be found in [67]:


With Chapter 4 the focus of the thesis shifts from perturbation theory to iterative methods for eigenvalue problems. The chapter is devoted to the analysis of the fairly recent and unobserved Subspace Projected Approximate Matrix (SPAM) [69] method. We analyze it from a numerical linear algebra perspective with particular interest for the choice of the approximating matrix. Some property preserving perturbations introduced in Chapter 2 will reappear here as desirable preconditioners not only theoretically but also numerically. At each point we illustrate the performance of SPAM in comparison with well-known methods such as Lanczos [49] or Jacobi-Davidson [72]. Sections 4.1 to 4.5 of this chapter can be found in [12]:


While an extended version of Section 4.6 will be found in [13]

1.2. ITERATIVE METHODS FOR EIGENVALUE PROBLEMS

Notation

Although most of the notation used is considered standard, it is useful to clarify a few assumptions. A more complete list of notational symbols and their meanings can be found on page 139.

When appropriate, we choose to label eigenvalues, but also singular values, in non-decreasing order of magnitude

\[ \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{n-1} \leq \lambda_n. \]

The spectrum of a matrix \( A \) is denoted by \( \sigma(A) \). Letters at the end of the alphabet represent vectors and these are always column vectors. Lower case Latin letters \( j, k \) and occasionally \( p \) and \( q \) are mostly used as indices. In this thesis, the letter \( i \) is used solely to indicate the imaginary unit.

The conjugate transpose of a matrix \( A \) is denoted by \( A^* \), also for \( A \) real. With \( I_n \) we mean the \( n \times n \) identity matrix and by \( e_j \) its \( j \)th column. Similarly, \( O_n \) stands for the zero matrix of size \( n \times n \) while \( \vec{0} \) stands for any of its columns. The commutator of \( A \) and \( B \) is denoted by \([A, B]\) and defined as \([A, B] = AB - BA\). Finally, the Hermitian and skew-Hermitian part of a matrix \( A \) are represented by \( \mathcal{H}(A) \) and \( \mathcal{S}(A) \) and defined as \( \mathcal{H}(A) = (1/2)(A + A^*) \) and \( \mathcal{S}(A) = (1/2)(A - A^*) \), respectively.