Efficient Methods For Nonlinear Eigenvalue Problems

Diploma Thesis

Timo Betcke

Technical University of Hamburg-Harburg
Department of Mathematics (Prof. Dr. H. Voß)

August 2002

Abstract

During the last years nonlinear eigenvalue problems of the type $T(\lambda)x = 0$ became more and more important in many applications with the rapid development of computing performance. In parallel the minmax theory for symmetric nonlinear eigenvalue problems was developed which reveals a strong connection to corresponding linear problems. We want to review the current theory for symmetric nonlinear eigenvalue problems and give an overview about present solution methods for arbitrary nonlinear eigenvalue problems. Based on the theory for symmetric problems we will introduce a new solution method for large sparse symmetric nonlinear eigenvalue problems which combines a Jacobi-Davidson type approach with the theory for symmetric problems. With an example from fluid-structure interaction we demonstrate the high performance of the new algorithm and give ideas how to extend this approach for nonsymmetric problems.

Contents

Chapter 1

Introduction

In this thesis we want to review methods for the nonlinear eigenvalue problem

$$T(\lambda)x = 0 \tag{1.1}$$

where $T(\lambda) \in \mathbb{C}^{n\times n}$ is a family of $n \times n$-matrices and $\lambda \in C$ is a complex parameter. A special case of this eigenvalue problem is the linear eigenvalue problem where

$$T(\lambda) = \lambda B - A.$$

In this case various methods are known for dense eigenvalue problems (e.g QR, Jacobi, Divide and Conquer, ...) and for large sparse problems (e.g. Arnoldi/Lanczos type methods, Jacobi-Davidson). Also the theory is at least for symmetric linear eigenvalue problems well understood today. Unfortunately the situation for nonlinear eigenvalue problems is not so clear. For dense problems there exist some methods based on Newton type approaches for finding solutions of (1.1). All these methods have the drawback that they need several LU or QR decompositons to find one solution of (1.1). Hence, they are not applicable for the computation of several eigenvalues of large sparse matrices since generally the computation time is too high. Therefore it is desirable to have projection methods for large sparse nonlinear eigenvalue problems in which the large problem is projected onto a suitable subspace from which approximations for the eigenvalues can be extracted. For general nonlinear eigenvalue problems Ruhe introduced an extension of the Rational Krylov method. Unfortunately, the convergence behaviour of this approach is not yet well understood. Up to now it seems to work well only for eigenvalue problems in which the nonlinearity has no great effects, for example in vibration analysis with small damping. Another drawback of this approach is that the structure of the eigenvalue problem is destroyed. For example symmetry properties in the eigenvalue problem are not preserved.

Our approach was to develop a projection method for nonlinear eigenvalue problems which preserves the structure of the problem. There we focused on symmetric problems. Symmetric nonlinear eigenvalue problem are an important class, since there

exist minmax-characterizations for the real eigenvalues of symmetric problems which reveal strong relationships to symmetric linear eigenvalue problems. By preserving the symmetry of the problem in a projection method we were able to exploit the minmax-characterizations to find an effective strategy to extract eigenvalues of the nonlinear problem. Our idea is based on a Jacobi-Davidson type approach which was already used by Sleijpen and van der Vorst for polynomial eigenvalue problems. But in our approach the projected problem is solved with a guarded iteration introduced by Werner in [38]. This ansatz allowed us to build up a subspace from which several eigenvalues of the nonlinear problem can be efficiently extracted. Another benefit is that decompositions of $T(\lambda)$ are not necessary any more to find eigenvalues of the nonlinear problem (1.1). The minmax-characterizations for nonlinear eigenvalue problems also allowed us to implement efficient restart strategies. So we were able to extract all 28 eigenvalues in a given interval from a rational eigenvalue problem of dimension 36040 with a maximum subspace of dimension 40. We will also demonstrate a successful preconditioning strategy which uses the idea to allow some few LU decompositions of $T(\lambda)$ if this is not too expensive. Other preconditioners like ILU or approximative inverses are certainly also possible. Finally we will present some possibilities to extend our approach for arbitrary nonlinear eigenvalue problems. The greatest difficulty is to guarantee that the algorithm does not converge towards already computed eigenvalues, since deflation strategies do not yet exist for nonlinear eigenvalue problems. In the symmetric case this is automatically guaranteed by our approach. Our results for the symmetric case can also be found in [3]. For the unsymmetric case we will give two ideas how to handle this problem and show an example eigenvalue problem where these strategies for unsymmetric problems already work well. But in the unsymmetric case there is still future research necessary.

The thesis is organized as follows. In the next chapter we will repeat some important facts for symmetric linear eigenvalue problems, which will be needed in later chapters. In chapter 3 we will discuss some iterative methods for linear eigenvalue problems. We will focus on inverse iteration, Arnoldi, Rational Krylov and Jacobi-Davidson. These methods will be the foundation of solution methods for nonlinear eigenvalue problems presented here. In chapter 4 we will discuss the minmax theory for symmetric nonlinear eigenvalue problems. Based on the analysis of quadratic problems we will introduce definitions for the Rayleigh functional and for overdamping and will introduce a numbering for eigenvalues of symmetric nonlinear problems. Based on this numbering the minmax theory for nonoverdamped eigenvalue problems developed by Voss and Werner is introduced. In chapter 5 we will then discuss solution methods for nonlinear eigenvalue problems. At first present methods for dense problems are discussed. Then the approach of Ruhe and Hager is presented. We will finally introduce the Jacobi-Davidson type approach for nonlinear eigenvalue problems. It is based on the minmax theory discussed in chapter 4 and uses the guarded iteration introduced in chapter 5 to solve the projected nonlinear problem in every step. In chapter 6 an example problem from fluid-solid vibration is used to demonstrate the speed of our new approach. Finally, in chapter 7 we will demonstrate some possibilities to adjust

our algorithm for nonsymmetric eigenvalue problems. But the ideas presented there are still subject of further research.

Throughout the thesis we will use capital letters for matrices, lower letters for vectors and greek letters for scalars. The complex conjugate transpose of a matrix A or vector x is denoted by A^H, respectively x^H. Lower indices at a matrix denote a principle submatrix. Hence, the matrix $H_{k+1,k}$ denotes the matrix $(h_{ij}), i = 1, \ldots, k+1, j = 1, \ldots, k$. The j-th column of a matrix A is denoted by a^j. An element of a vector x is denoted with $x(j)$. Hence, $x(1)$ is the first element of the vector x. The matrix I always denotes the identity matrix and the vector e_j denotes the j-th unity vector. For the terminology of linear eigenvalue problems we use the following usual definition. If we are speaking of an eigenvector $\mathbf{x}$ of a matrix A we always mean right eigenvectors, hence

$$Ax = \lambda x$$

for the corresponding eigenvalue λ. If a left eigenvector $\mathbf{y}$ of A is meant it will be emphasized in the text.

With $\langle x, y \rangle$ we always denote the usual scalar product

$$\langle x, y \rangle = \sum_{i=1}^{n} x(i)\overline{y(i)}$$

if not otherwise stated in the text. The norm $\|\cdot\|$ shall always denote the 2-norm of a matrix or vector.

Chapter 2

Hermitian linear eigenvalue problems

Now we want to repeat some basic facts for Hermitian eigenvalue problems as they will be needed in subsequent chapters.

Let us consider the Hermitian eigenvalue problem

$$Ax = \lambda x \tag{2.1}$$

with $A \in \mathbb{C}^{n\times n}$, $A = A^H$. Obviously A is normal. Therefore A is diagonalizable with a unitary matrix U and we have $U^H AU = \Lambda$. It is easy to show that all eigenvalues of A are real. Thus we can order the eigenvalues according to

$$\lambda_{min} = \lambda_1 \leq \lambda_2 \cdots \leq \lambda_{n-1} \leq \lambda_n = \lambda_{max}$$

The eigenvalues of A can be characterized with the Rayleigh quotient, which is defined as

$$\rho(x) := \frac{x^H Ax}{x^H x}, \quad x \neq 0$$

For Hermitian matrices $\rho(x)$ is always real since

$$x^H Ax = x^H A^H x = (Ax)^H x = \overline{x^H Ax}.$$

Another important fact of the Rayleigh quotient is that the eigenvectors are stationary points of $\rho(x)$, because

$$\nabla\rho(x) = \frac{2}{x^H x}\left[Ax - \rho(x)x\right].$$

The Rayleigh quotient can be used to classify Hermitian matrices.

DEFINITION 2.1
A Hermitian matrix A is said to be

- *positive (negative) definite,*
  if $\rho(x) > 0$ ($\rho(x) < 0$) for all $x \in \mathbb{C}^n$,
- *positive (negative) semi-definite,*
  if $\rho(x) \geq 0$ ($\rho(x) \leq 0$) for all $x \in \mathbb{C}^n$,
- *indefinite,*
  otherwise.

THEOREM 2.2
For a Hermitian matrix $A \in \mathbb{C}^{n\times n}$ the following properties are equivalent:

(i) *A is positive definite,*

(ii) *All eigenvalues of A are positive,*

(iii) *Let $A_i := \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1i} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix}$, then*
$detA_i > 0$ for all $i = 1, \dots, n$,

(iv) *A has a unique Cholesky factor C which is lower triangular with positive diagonal and satisfies $A = CC^H$.*

Proof:
[12]

The following theorem gives a first characterization for the eigenvalues of A with the Rayleigh quotient $\rho(x)$.

THEOREM 2.3
(Rayleigh)

(i)
$$\lambda_1 \leq \rho(x) \leq \lambda_n \qquad \forall x \in \mathbb{C}^n \backslash \{0\}$$

(ii)
$$\lambda_1 = \min_{x\neq 0} \rho(x), \quad \lambda_n = \max_{x\neq 0} \rho(x)$$

(iii) *If $x \neq 0$ and $\lambda_1 = \rho(x)$ resp. $\lambda_n = \rho(x)$, then x is eigenvector of λ_1 resp. λ_n.*

(iv)

$$\begin{aligned}\lambda_i &= \min\{\rho(x) : x^H u^j = 0, j = 1, \dots i-1, x \neq 0\} \\ &= \max\{\rho(x) : x^H u^j = 0, j = i+1, \dots n, x \neq 0\}\end{aligned}$$

Proof:

(i) A is Hermitian. Thus there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ with $A = U\Lambda U^H$, $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Then with $\tilde{x} := x/ \parallel x \parallel$ we have

$$\rho(x) = \frac{x^H A x}{x^H x} = (U^H\tilde{x})^H \Lambda (U^H\tilde{x}) = \sum_{k=1}^{n} \lambda_k |(U^H\tilde{x})(k)|^2 \tag{2.2}$$

We can conclude

$$\lambda_1 \sum_{k=1}^{n} |(U^H\tilde{x})(k)|^2 \leq \sum_{k=1}^{n} \lambda_k |(U^H\tilde{x})(k)|^2 \leq \lambda_n \sum_{k=1}^{n} |(U^H\tilde{x})(k)|^2. \tag{2.3}$$

Since U is unitary,

$$\sum_{k=1}^{n} |(U^H\tilde{x})(k)|^2 = 1.$$

So it follows from (2.3) that

$$\lambda_1 \leq \sum_{k=1}^{n} \lambda_k |(U^H\tilde{x})(k)|^2 \leq \lambda_n.$$

(ii) We just have to show that the inequalities in *(i)* are sharp. If x is eigenvector to λ_1, respectively λ_n we have $\rho(x) = \lambda_1$, respectively $\rho(x) = \lambda_n$.

(iii) If $\lambda_1 = \rho(x)$ and r is the multiplicity of the eigenvalue λ_1, then $(U^H\tilde{x})(k)$ must be zero for $k = r+1, \dots, n$. Thus x is eigenvector to λ_1. For λ_n an analogue argumentation holds.

(iv) If $x \perp u^j, j = 1, \dots, i-1$ then

$$\rho(x) = \sum_{k=i}^{n} \lambda_k |(U^H\tilde{x})(k)|^2$$

and we can conclude as in *(i)*. A similar argumentation holds for $x \perp u^j, j = i+1, \dots, n$.

Another characterization which does not depend on the eigenvectors of A is the "maxmin theorem" of Courant-Fischer.

THEOREM 2.4
(Courant-Fischer)

$$\begin{aligned}\lambda_i &= \max_{\{p_1,\ldots,p_{i-1}\}} \min\{\rho(x) : x \neq 0, x^H p_j = 0, j = 1,\ldots,i-1\} \\ &= \min_{\{p_1,\ldots,p_{n-i}\}} \max\{\rho(x) : x \neq 0, x^H p_j = 0, j = 1,\ldots,n-i\}\end{aligned}$$

Proof:
Let

$$h(p_1,\ldots,p_{i-1}) := \min\{\rho(x) : x \neq 0, x^H p_j = 0, j = 1,\ldots,i-1\}.$$

Then with the Rayleigh-Theorem we have

$$h(u^1, u^2, \ldots, u^{i-1}) = \lambda_i.$$

For arbitrary $p_1,\ldots,p_{i-1} \in \mathbb{C}^n$ there is always an $x \neq 0$, which fulfills the equations

$$x^H p_j = 0, j = 1,\ldots,i-1, \quad x^H u^j = 0, j = i+1,\ldots,n$$

From the last equation and the Rayleigh-Theorem we know that $\rho(x) \leq \lambda_i$. Thus we have $h(p_1,\ldots,p_{i-1}) \leq \lambda_i$.
The prove of the second formulation for eigenvalues is similar.

We will especially need the following characterization in subsequent chapters:

THEOREM 2.5
(Poincaré)

$$\lambda_i = \min_{dimV=i} \max_{x\in V\setminus\{0\}} \rho(x) = \max_{dimV=n-i+1} \min_{x\in V\setminus\{0\}} \rho(x)$$

Proof:
The maximum in the second formulation of the Courant-Fischer-Theorem is achieved by the eigenvectors $u^{i+1},\ldots,u^n$. For these vectors the Courant-Fischer-Theorem results in

$$\lambda_i = \max_{x\in \mathrm{span}\{u^1,\ldots,u^i\}} \rho(x)$$

By restricting the vectors $p^1,\ldots,p^n - i$ to linear independent set of vectors in the Courant-Fischer-Theorem we finally arrive at the full formulation of the Poincaré-Theorem.

$$\lambda_i = \min_{\dim V=i} \max_{x\in V\setminus\{0\}} \rho(x)$$

The proof for the second formulation of the theorem is similar. The minmax characterizations from Courant-Fischer and Poncaré in this chapter are also valid if we consider the generalized Hermitian eigenvalue problem

$$Ax = \lambda Bx,$$

where A is Hermitian and B Hermitian positive definite. Then the Rayleigh quotient $\rho(x)$ has to be replaced with

$$\tilde{\rho}(x) = \frac{x^H A x}{x^H B x}$$

Since the eigenvectors are orthogonal to the B-inner product in the generalized Hermitian eigenvalue problem, the $B-$ inner product $\langle x, By \rangle$ in the Courant-Fischer theorem has to be used.

Chapter 3

Methods for linear eigenvalue problems

In this chapter we want to review some methods for the linear eigenvalue problem

$$Ax = \lambda x, \tag{3.1}$$

which will be useful for the treatment of nonlinear eigenvalue problems in later chapters. All methods for problem (3.1) must be in some way iterative because the eigenvalues of A are the roots of its characteristic polynomial which in general can not be exactly determined with a finite number of elementary operations.

Numerical methods for eigenvalue problems can roughly be distinguished into methods for the full eigenvalue problem and into methods which compute some eigenvalues and eigenvectors. In the first category there are for example QR, Jacobi, Divide and Conquer or Twisted Factorizations. QR and Jacobi are based on the idea to perform similarity transformations with orthogonal matrices of the form $A^{(i+1)} = {Q^{(i)}}^H A^{(i)} Q^{(i)}$ which converge against the Schur-Form $Q^H A Q = R$ of A. Divide-and-Conquer follows a recursive approach to compute the eigenvalues and eigenvectors of a symmetric matrix and Twisted-Factorizations is based on a careful implementation of inverse iteration for computing the eigenvectors of symmetric tridiagonal matrices. These methods neeed up to $O(n^3)$ Flops (Twisted Factorizations needs $O(n^2)$ flops) for an eigenvalue problem of dimension n. This is sufficient for problems of small dimension where *small* certainly depends on the performance of the computer. Usually it is up to some hundred variables.

For large dimensions the full linear eigenvalue problem can normally not be solved in a reasonable amount of time. For many applications this is not necessary. Very often it is sufficient to only know the eigenvalues in a certain range of the spectrum. For these types of eigenvalue problems there are several different approaches which deal with different type of problems. Here we will review inverse iteration, Arnoldi,

Rational Krylov and Jacobi-Davidson. These methods will lead to ideas how to solve the nonlinear eigenvalue problem which is later discussed in this thesis.

3.1 Inverse iteration

Inverse iteration is a fast method for computing one eigenvalue and the corresponding eigenvector near a given shift. It is derived from the power method for computing the eigenvalue with largest modulus and a corresponding eigenvector. Let us consider the linear eigenvalue problem (3.1). Then the power method is defined as:

Algorithm 1 The Power Method

Start with a nonzero initial vector v_0.
for $k = 1, 2, \ldots$ until convergence **do**

$$v_k = \frac{1}{\alpha_k} A v_{k-1}$$

where α_k is a convenient scaling factor.
end for

THEOREM 3.1
Assume that there is one and only one eigenvalue λ_1 of largest modulus and that the algebraic multiplicity of λ_1 equals its geometric multiplicity. Then either the initial vector v_0 has no component in the invariant subspace associated with λ_1 or the sequence of vectors generated by Algorithm 1 converges to an eigenvector associated with λ_1.

Proof:
[28]

The power method converges linearly to an eigenvector according to the eigenvalue with largest modulus. If we order the eigenvalues of A according to their modulus

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|,$$

then the convergence factor of the method is given by

$$\rho = \frac{|\lambda_2|}{|\lambda_1|}$$

Therefore convergence of the method may be very slow if the modulus of λ_2 is close to the modulus of λ_1. To solve this problem and to generalize the power method for the computation of arbitrary eigenvalues, we consider the shifted problem

$$\frac{x}{\lambda - \sigma} = (A - \sigma I)^{-1} x, \tag{3.2}$$

where $\sigma \in \mathbb{C}$ is a given shift. If Algorithm 1 is applied to $(A - \sigma I)^{-1}$ we achieve convergence to the eigenvector corresponding to the eigenvalue which is nearest to the shift σ. This is the basic form of inverse iteration with fixed shifts. The speed of convergence to the eigenvalue λ_j nearest to σ is determined from

$$\rho = \max_{i=1,\ldots,n, i \neq j} \frac{|\sigma - \lambda_j|}{|\sigma - \lambda_i|}.$$

Inverse iteration converges very fast if σ is a good approximation of λ_j. With a slight modification of the inverse iteration we can finally arrive at a method that has in general quadratic convergence and for Hermitian matrices even cubic convergence.

In chapter 2 it was already mentioned that the eigenvectors of a Hermitian matrix are stationary points of the corresponding Rayleigh quotient. If $(\hat{\lambda}, \hat{x})$ is an eigenpair of a A and A is Hermitian, the Rayleigh quotient of a slightly deturbed vector $x = \hat{x} + \epsilon$ equals

$$\rho(x) = \rho(\hat{x} + \epsilon) = \rho(x) + O(\|\epsilon\|^2) = \hat{\lambda} + O(\|\epsilon\|^2).$$

Thus a perturbation of order $\|\epsilon\|$ in the eigenvector approximation leads to a perturbation of $\|\epsilon\|^2$ in the approximation of the eigenvalue. If A is not Hermitian it is still advisable to use the Rayleigh quotient because

$$\min_{\mu \in \mathbb{C}} \frac{\|Ax - \mu x\|}{\|x\|} = \frac{x^H A x}{x^H x}.$$

for arbitrary $A \in \mathbb{C}^{n \times n}, x \in \mathbb{C}^n$. This features of the Rayleigh quotient can be used to update the shift σ in the inverse iteration in every step with the Rayleigh quotient of the vector v_k just computed. This is described in the following algorithm.

Algorithm 2 Inverse Iteration with Rayleigh-Shifts

Start with a nonzero initial vector v_0 and an initial shift σ_0.

for $k = 1, 2, \ldots$ until convergence **do**

$$(A - \sigma_{k-1} I) v_k = \frac{1}{\alpha_{k-1}} v_{k-1}$$

where α_{k-1} is a convenient scaling factor.

$$\sigma_k = \frac{v_k^H A v_k}{v_k^H v_k}$$

end for

Now we want to discuss the speed of convergence for Algorithm 2 more in detail (cf. [33]). For the discussion consider that the sequence of vectors converges to an eigenvector corresponding to a simple eigenvalue of A. Let R be a unitary matrix with

$y_k = R^H v_k$ then

$$\sigma_k = \frac{y_k^H R^H A R y_k}{y_k^H y_k}$$

and

$$(R^H A R - \sigma_k I) y_{k+1} = \frac{1}{\alpha_k} y_k.$$

If we assume that $\|v_k\| = 1$ and $R = (v_k, U)$, where U is chosen such that R is unitary. Then $y_k = e_1$ and

$$R^H A R = \begin{pmatrix} \sigma_k & h^H \\ g & C \end{pmatrix},$$

Therefore y_{k+1} can be determined from the solution of the system

$$\begin{pmatrix} 0 & h^H \\ g & C - \sigma_k I \end{pmatrix} y_{k+1} = \frac{1}{\alpha_k} e_1.$$

If we choose α_k such that the first component of y_{k+1} is unity and write y_{k+1} in the form $y_{k+1} = (1, p^T)^T$, then

$$p = -(C - \sigma_k I)^{-1} g.$$

The residual of y_k as eigenvector approximation of $R^H A R$ is defined as

$$r_k := \frac{R^H A R y_k - \sigma_k y_k}{\|y_k\|}.$$

Now we want to establish a correspondence between $\|r_k\|$ and $\|r_{k+1}\|$. Since $y_k = e_1$ it is easily determined that

$$\|r_k\| = \|g\|.$$

Now suppose that Q is a unitary matrix such that $Q^H y_{k+1} = \|y_{k+1}\| e_1$. Then if

$$Q^H R^H A R Q = \begin{pmatrix} \sigma_{k+1} & \tilde{h}^H \\ \tilde{g} & \tilde{C} \end{pmatrix}$$

we have

$$\|r_{k+1}\| = \|\tilde{g}\|.$$

The unitary matrix Q can be constructed by

$$Q = \begin{pmatrix} 1 & -p^H \\ p & I \end{pmatrix} D.$$

Let

$$\begin{pmatrix} -p^H \\ I \end{pmatrix} = \tilde{Q} R,$$

with $\tilde{Q}^H \tilde{Q} = I_{n-1}$ and R upper triangular with $R \in \mathbb{C}^{n-1 \times n-1}$. Then D can be defined as

$$D = \begin{pmatrix} (1 + \|p\|^2)^{-1/2} & 0 \\ 0 & R^{-1} \end{pmatrix}.$$

A short calculation shows that Q is unitary and $Q^H y_{k+1} = \|y_{k+1}\| e_1$. For $\tilde{g}$ it follows that

$$\tilde{g} = R^{-H}(1 + \|p\|^2)^{-1/2}(-\sigma_k p - p h^H p + g + Cp)$$

With the equality $(C - \sigma_k I)p = -g$ this can be simplified to

$$\tilde{g} = -R^{-H}(1 + \|p\|^2)^{-1/2} p h^H p.$$

Thus, we have

$$\begin{aligned} \|r_{k+1}\| &\leq (1 + \|p\|)^{-1/2} \|R^{-H}\| \|p\|^2 \|h\| \\ &\leq (1 + \|p\|)^{-1/2} \|R^{-H}\| \|(C - \sigma_k I)^{-1}\|^2 \|h\| \|g\|^2 \\ &= (1 + \|p\|)^{-1/2} \|R^{-H}\| \kappa \|r_k\|^2 \end{aligned} \tag{3.3}$$

where

$$\kappa = \|(C - \sigma_k I)^{-1}\|^2 \|h\|.$$

If p is zero, then $R = I_{n-1}$. Hence, if v_k is near an eigenvector, then p is almost zero and the factor $(1 + \|p\|)^{-1/2}\|R^{-H}\|$ is near 1.

To understand the influence of κ we introduce the spectral condition number of the eigendirection span(x) (cf. [4]).

Definition 3.2 (Spectral condition number of an eigendirection)
Let λ be a simple eigenvalue of A, x a corresponding eigenvector, $Q \in \mathbb{C}^{n \times n-1}$ chosen such that $\{x, q^1, q^2, \ldots, q^{n-1}\}$ is a unitary basis of $\mathbb{C}^n$ and $B = Q^H A Q$ the projection of A onto the space spanned by Q. Then the spectral condition number csp(x) of the eigendirection span(x) is defined by

$$csp(x) = \|Q(B - \lambda I)^{-1} Q^H\|.$$

If (σ_k, v_k) is a good approximation to the eigenpair (λ, x) of A,
the value $\|(C - \sigma_k I)^{-1}\|$ is near to the spectral condition csp(x) . Hence

$$\kappa \approx \mathrm{csp}(x)^2 \|h\|.$$

For the analysis it was necessary to assume that Algorithm 2 converges against an eigenvector to a simple eigenvalue to ensure that $(C - \sigma_k I)^{-1}$ exists near the wanted eigenvalue. Equation (3.3) shows the quadratic convergence near a simple eigenvalue. If the matrix A is Hermitian we have $g = h$ which leads to

$$\|r_{k+1}\| \lessapprox (1 + \|p\|)^{-1/2} \|R^{-H}\| \ \mathrm{csp}(x)^2 \|r_k\|^3,$$

showing the cubic convergence of inverse iteration for a Hermitian matrix.

In every step of Algorithm 2 the linear system

$$(A - \sigma_k I) v_{k+1} = \frac{1}{\alpha_k} v_k \tag{3.4}$$

is solved. If σ_k converges towards an eigenvalue of A, the matrix $(A-\sigma_k I)$ is becoming more and more ill-conditioned. Hence a large error in the solution of (3.4) can be expected. Nevertheless, this error does not negatively influence the convergence of Algorithm 2.

THEOREM 3.3
Let λ be a simple eigenvalue of A and x the corresponding eigenvector. If the vector x is well-conditioned, the error made in solving (3.4) is mainly in the direction generated by x, which is the direction required.

Proof:
[4]

In chapter 5.1.1 we will present generalizations of inverse iteration for nonlinear eigenvalue problems which also reach cubic convergence in the symmetric case. Inverse iteration with Rayleigh shifts heavily depends on efficient solvers for the linear system (3.4) because in every step of Algorithm 2 a new LU decomposition is necessary.

In the next sections we will discuss an approach which delivers approximations to several eigenvalues simultaneously without needing to perform LU decompositions. But if only very view eigenvalues are wanted and linear systems with A can be efficiently solved, inverse iteration delivers very swift convergence.

3.2 Arnoldi's Method for linear eigenvalue problems

In this section we want to review Arnoldi's method for linear eigenvalue problems. It was originally introduced in 1951 as procedure to reduce matrices to Hessenberg form. But as truncated procedure Arnoldi's method leads to a good technique for approximating the eigenvalues of a matrix. It can be understood as projection procedure onto inflating subspaces. The great advantage compared to inverse iteration is that only matrix-vector multiplications need to be performed in the basic algorithm. Large sparse matrices which are typical in many applications often allow to perform this operation with the complexity $O(n)$.

To introduce the method, we first collect some properties for Krylov subspaces.

DEFINITION 3.4
Let $v \in \mathbb{C}^n$, $A \in \mathbb{C}^{n\times n}$, then the Krylov subspace $K_m(A,v)$ is defined as

$$K_m(A,v) \equiv span\{v, Av, A^2v, \ldots, A^{m-1}v\}$$

If the context is clear we will just write K_m instead of $K_m(A,v)$.

The elements of K_m can be written in terms of polynomials of A.

PROPOSITION 3.5
The Krylov subspace K_m is the subspace of all vectors in $\mathbb{C}^n$ which can be written as $x = p(A)v$, where p is a polynomial of degree not exceeding $m-1$.

The dimension of K_m can be characterized with the following two propositions (cf. [28])

PROPOSITION 3.6
Let μ be the degree of the minimal polynomial of v. Then K_μ is invariant under A and $K_m = K_\mu$ for all $m \geq \mu$.

PROPOSITION 3.7
The Krylov subspace K_m is of dimension m if and only if the degree of the minimal polynomial of v with respect to A is larger then $m-1$.

Now we can introduce Arnoldi's algorithm which will prove to be a method for projecting a matrix A onto the Krylov subspace K_m.

Algorithm 3 Basic Arnoldi iteration

Choose a vector v_1 with $\|v_1\| = 1$.
for $j = 1, \ldots, m$ **do**
  $h_{ij} = v_i^H A v_j$, $i = 1, 2, \ldots, j$
  $w_j = Av_j - \sum_{i=1}^{j} h_{ij} v_i$
  $h_{j+1,j} = \|w_j\|$, if $h_{j+1,j} = 0$ stop
  $v_{j+1} = w_j / h_{j+1,j}$
end for

PROPOSITION 3.8
The vectors $v_1, v_2, \ldots, v_m$ form an orthonormal basis of the subspace $K_m(A, v)$.

Proof:
The vectors $v_j, j = 1, \ldots, m$ are orthonormal by construction. We will show by induction on j that each vector v_j is of the form $v_j = q_{j-1}(A)v_1$ where q_{j-1} is a polynomial of degree $j-1$. We have $v_1 = q_0(A)v_1$ with $q_0(A) \equiv 1$. Assume that the result holds for all $i \leq j$. Then

$$h_{j+1,j} v_{j+1} = Av_j - \sum_{i=1}^{j} h_{ij} v_i = A q_{j-1}(A) v_1 - \sum_{i=1}^{j} h_{ij} q_{i-1}(A) v_1.$$

Hence v_{j+1} can be written as $q_j(A)v_1$.

In every step of Algorithm 3 we have the recursion

$$Av_j = \sum_{i=1}^{j} h_{ij} v_i + h_{j+1,j} v_{j+1}$$

Collecting all steps of the Algorithm into one equation yields

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^H, \tag{3.5}$$

where V_m denotes the $n \times m$ matrix with column vectors $v_1, \ldots, v_m$ and H_m denotes the $m \times m$ Hessenberg matrix whose entries are defined by Algorithm 3. By premultiplying the last equation with V_m we arrive at

$$V_m^H A V_m = H_m.$$

Hence, the matrix H_m is the orthogonal projection of A onto K_m. The strategy is now to compute the eigenpairs $(\hat{\lambda}_i, \hat{x}_i), i = 1, \ldots, m$ of H_m. The values $\hat{\lambda}_i$ are called Ritz values, the corresponding approximations for eigenvectors $V_m \hat{x}_i$ are called Ritz vectors. If A is Hermitian the Hessenberg matrix H_m reduces to a tridiagonal matrix and the recursion (3.5) becomes a three-term recurrence.

Arnoldi's method stops if $h_{j+1,j}$ is zero. This is called a 'lucky breakdown'. The next proposition gives a statement when this situation occurs.

PROPOSITION 3.9
Arnoldi's algorithm breaks down at step j (i.e., $h_{j+1,j} = 0$) if and only if the minimal polynomial of v_1 is of degree j. Moreover, in this case the subspace $K_j(A, v)$ is invariant and the approximate eigenvalues and eigenvectors are exact.

Proof:
[28]

The residual of the eigenvalue and eigenvector approximations can be directly computed from the projected problem.

THEOREM 3.10
Let $y_i^{(m)}$ be an eigenvector of H_m associated with the eigenvalue $\lambda_i^{(m)}$ and $u_i^{(m)}$ the Ritz approximate eigenvector $u_i^{(m)} = V_m y_i^{(m)}$. Then,

$$(A - \lambda_i^{(m)} I) u_i^{(m)} = h_{m+1,m} e_m^H y_i^{(m)} v_{m+1}$$

and, therefore,

$$\|(A - \lambda_i^{(m)} I) u_i^{(m)}\| = h_{m+1,m} |e_m^H y_i^{(m)}|.$$

Proof:
The theorem follows directly from equation (3.5) by multiplying both sides with $y_i^{(m)}$.

$$\begin{aligned} AV_m y_i^{(m)} &= V_m H_m y_i^{(m)} + h_{m+1,m} e_m^H y_i^{(m)} v_{m+1} \\ &= \lambda_i^{(m)} V_m y_i^{(m)} + h_{m+1,m} e_m^H y_i^{(m)} v_{m+1} \end{aligned}$$

Hence,

$$AV_m y_i^{(m)} - \lambda_i^{(m)} V_m y_i^{(m)} = h_{m+1,m} e_m^H y_i^{(m)} v_{m+1}.$$

In the way Arnoldi's method is presented here, it is not used in practical implementations. There are several issues that have to be further investigated to make Arnoldi's method efficient. So all old basis vectors must be stored if the matrix A is not Hermitian. Therefore restart techniques are needed to ensure that the size of the basis stays small. Deflation if an eigenvalue is found is also an important topic. In practical implementations of the algorithm the finite precision arithmetic of modern computers also has to be considered. For example the orthogonalization in every step must be carefully implemented to achieve numerically orthogonal eigenvectors. A good starting point for these topics is [28].

In Arnoldi's Method the extremal eigenvalues converge first while the convergence of interior eigenvalues is often extremely slow. A possibility to overcome this drawback is to consider a shifted and inverted problem as it is done in the inverse iteration. Let us consider the generalized eigenvalue problem

$$Ax = \lambda Bx.$$

Then a strategy is to consider the shifted and inverted eigenvalue problem

$$(A - \sigma B)^{-1} Bx = \frac{1}{\lambda - \sigma} x$$

and to apply Arnoldi's method to this new problem. The eigenvalues near the shift converge fast as they are now extremal eigenvalues. More information to this Shift-And-Invert-Arnoldi can be found in [15]. In the next section we want to review a method to apply several shifts in one Arnoldi run. This allows the computation of eigenvalues in several regions of the spectrum by incorporating several shifts in interesting areas of the spectrum into one Arnoldi basis.

3.3 The Rational Krylov method for regular pencils

In this section we want to find eigenpairs of the regular pencil

$$(A - \lambda B)x = 0 \tag{3.6}$$

The Rational Krylov method starts as Shift-And-Invert-Arnoldi method with shift σ_1. After j steps we obtain the basic recursion

$$(A - \sigma_1 B)^{-1} B V_j = V_{j+1} H_{j+1,j}, \tag{3.7}$$

where V_j is the orthogonal basis of the Krylov-Subspace in step j. After a moderate number of steps j we obtain with the solution of the eigenvalue problem

$$H_{j,j} Z^{(j)} = Z^{(j)} \Theta^{(j)}$$

the Ritz-Values $\lambda_i^{(j)} = \sigma_1 + \frac{1}{\theta_i^{(j)}}$ which are a good approximation for eigenvalues of problem 3.6 near the shift σ_1. However, when the spectrum of 3.6 near σ_1 is fully exploited, one often wishes to find eigenvalues near a second shift σ_2 without throwing away the information gathered at the old shift σ_1. So our goal is to find a new formulation of the basic recursion 3.7 with the new shift σ_2. The recursion 3.7 can be rewritten as

$$(\sigma_1 - \sigma_2) B V_{j+1} H_{j+1,j} + B V_j = (A - \sigma_2 B) V_{j+1} H_{j+1,j}. \tag{3.8}$$

This leads to

$$B V_{j+1} (I_{j+1,j} + (\sigma_2 - \sigma_1) H_{j+1,j}) = (A - \sigma_2 B) V_{j+1} H_{j+1,j} \tag{3.9}$$

Define the new matrix $K_{j+1,j}$ as $K_{j+1,j} := I_{j+1,j} + (\sigma_2 - \sigma_1) H_{j+1,j}$. By multiplying the last equation from the left with $(A - \sigma_2 B)^{-1}$ we finally obtain the equation

$$(A - \sigma_2 B)^{-1} B V_{j+1} K_{j+1,j} = V_{j+1} H_{j+1,j}. \tag{3.10}$$

This is already an equation similar to the basic recursion 3.7. The last step to receive an Arnoldi recursion is to get rid of the matrix $K_{j+1,j}$ on the left hand side of the equation. To do this, first perform a QR-decomposition of $K_{j+1,j} = Q_{j+1,j+1} \begin{bmatrix} R_{j,j} \\ 0 \end{bmatrix}$. It is easily shown that the matrix $R_{j,j}$ is regular if the Hessenberg matrix $H_{j+1,j}$ is unreduced. So we can insert the QR decomposition of $K_{j+1,j}$ into the recursion and multiply from the right with the inverse of $R_{j,j}$ to obtain

$$(A - \sigma_2 B)^{-1} B V_{j+1} Q_{j+1,j} = V_{j+1} H_{j+1,j} R_{j,j}^{-1}. \tag{3.11}$$

With the new basis $\tilde{W}_j := V_{j+1} Q_{j+1,j}$ and $L_{j+1,j} := Q_{j+1,j+1}^H H_{j+1,j} R_{j,j}^{-1}$ the last equation can be written as

$$(A - \sigma_2 B)^{-1} B \tilde{W}_j = \tilde{W}_{j+1} L_{j+1,j} \tag{3.12}$$

This is almost the desired Arnoldi recursion. Only $L_{j+1,j}$ needs to be transformed to Hessenberg form. By performing a bottom-up Hessenberg transformation (cf. [25]) $L_{j+1,j}$ can be decomposed into Hessenberg form with

$$L_{j+1,j} = \begin{bmatrix} P_j & 0 \\ 0 & 1 \end{bmatrix} \tilde{H}_{j+1,j} P_j^H. \tag{3.13}$$

By multiplying 3.12 from the right with P_j we receive

$$(A - \sigma_2 B)^{-1} B \tilde{W}_j P_j = \tilde{W}_{j+1} \begin{bmatrix} P_j & 0 \\ 0 & 1 \end{bmatrix} \tilde{H}_{j+1,j}. \tag{3.14}$$

With $W_{j+1} := \begin{bmatrix} \tilde{W}_j P_j & \tilde{w}_{j+1} \end{bmatrix}$ we receive the Arnoldi recursion

$$(A - \sigma_2 B)^{-1} B W_j = W_{j+1} \tilde{H}_{j+1,j}. \tag{3.15}$$

This recursion is similar to the recursion 3.7 with the difference that we have a recursion for the shift σ_2 without losing information from the old basis V_j. In an actual implementation one would iterate with a fixed-shift σ_1 until a sufficient number of eigenvalues near σ_1 have converged. Then the shift is changed to σ_2 to lead the recursion towards eigenvalues near σ_2. In addition locking and purging can be applied to control the dimension of the basis V. Algorithm 4 summarizes the Rational Krylov Algorithm for Generalized Hermitian eigenvalue problems. Further details can be found in ([1],[22],[23],[26]).

Algorithm 4 Rational Krylov Subspace Method for Generalized Hermitian Eigenvalue Problems

1: **s**tart with σ_1 starting shift, v_1 unit starting vector, basis size $j = 1$
2: **for** $i = 1, 2, \ldots$ **until** convergence **do**
3: expand j-step Arnoldi recursion to m steps:
$(A - \sigma_i B)^{-1} B V_m = V_{m+1} H_{m+1,m}$
4: compute Ritz values, *lock*, and *purge*
5: determine new shift σ_{i+1} in interesting region
6: factorize $I_{j+1,j} + (\sigma_{i+1} - \sigma_i) H_{j+1,j} = Q_{j+1,j+1} R_{j+1,j}$
7: $H_{j+1,j} := \begin{bmatrix} P_j^* & 0 \\ 0 & 1 \end{bmatrix} Q^*_{j+1,j+1} H_{j+1,j} R^{-1}_{j,j} P_j$
8: $V_{j+1} := V_{j+1} Q_{j+1,j+1} \begin{bmatrix} P_j^* & 0 \\ 0 & 1 \end{bmatrix}$
9: **end for**

3.4 Jacobi-Davidson

The Jacobi-Davidson method introduced by Sleijpen and van der Vorst (cf. [29],[31],[32]) is an orthogonal projection method for (3.1). Hence, if V is a matrix with orthonormal columns we extract approximations for the eigenvalues of A from the projected eigenvalue problem

$$V^H A V y = \sigma y.$$

The corresponding Ritz pair (σ, u), $u = Vy$ approximates an eigenpair of A. In order to improve the approximation the subspace V must be extended in a certain direction. The most desirable orthogonal correction for u fulfills the equation

$$A(u + t) = \lambda(u + t), \quad t \perp u \tag{3.16}$$

Because we are seeking for a correction orthogonal to u we restrict the operator A to the subspace orthogonal to u. Let

$$\tilde{A} := (I - uu^H)A(I - uu^H).$$

Then A can be written as

$$A = \tilde{A} + Auu^H + uu^H A - u^H Au\, uu^H. \tag{3.17}$$

Now let $r := Au - \sigma u$, $\sigma = u^H Au$ and $\|u\| = 1$. By inserting (3.17) into (3.16) we arrive at

$$\tilde{A}t - \lambda t = -r + (\lambda - \sigma - u^H At)u$$

The vector t on the left side is mapped onto $\{u\}^\perp$. The residual r on the right side is orthogonal to u via the the Ritz-Galerkin condition for the projected problem and t itself is orthogonal to u from the definition of the correction t. Hence, multiplying the last equation with u^H leads to

$$(\lambda - \sigma - u^H At) = 0,$$

and therefore

$$\tilde{A}t - \lambda t = -r.$$

Since $t \perp u$ the last equation can be written as

$$(I - uu^H)(A - \lambda I)(I - uu^H)t = -r, \quad t \perp u.$$

Unfortunately we do not know λ. But we have the approximaton σ for λ. Hence, substituting λ with σ in the last equation, the Jacobi-Davidson correction equation is finally obtained as

$$(I - uu^H)(A - \sigma I)(I - uu^H)t = -r, \quad t \perp u. \tag{3.18}$$

A basic framework for the Jacobi-Davidson method can be written in the following form.

Algorithm 5 Jacobi-Davidson for linear eigenvalue problems

1: Start with $V = v_1/\|v_1\|$
2: $n = 1$, $k = 1$
3: **while** $n \leq$ Number of wanted Eigenvalues **do**
4: Compute the wanted eigenvalue θ and the corresponding eigenvector y of $V^H AVy = \theta y$.
5: $\sigma_k = \theta$, $u_k = Vy$, $r_k = (Au_k - \sigma_k u_k)/\|u_k\|$
6: **if** $\|r_k\| < \epsilon$ **then**
7: PRINT σ_k,u_k
8: $n = n + 1$
9: Apply a Deflation Strategy to remove u_k from V.
10: GOTO 3
11: **end if**
12: Find an approximate solution for the correction equation

$$(I - \frac{u_k u_k^H}{u_k^H u_k})(A - \sigma_k I)(I - \frac{u_k u_k^H}{u_k^H u_k})t = -r_k.$$

13: $t = t - VV^T t$,$\tilde{v} = t/\|t\|$, $V = [V, \tilde{v}]$
14: If necessary perform a purge-operation to reduce the size of the subspace V.
15: $k = k + 1$
16: **end while**

Simple deflation and restart strategies for Jacobi-Davidson are mentioned in ([29]).

There are several ways to interpret the correction equation, which shall be discussed now (cf. [32]).

Equation (3.18) can be written as

$$(A - \sigma I)t - \alpha u = -r, \quad t \perp u,$$

where $\alpha \in \mathbb{C}^n$ is chosen such that $t \perp u$. From now on $A - \sigma I$ shall be approximated by a matrix M, for which linear equations can easily be solved. This leads to approximations $\tilde{t}$ for t for which the following holds:

$$\begin{aligned} M\tilde{t} - \alpha u &= -r, \\ \tilde{t} &= \alpha M^{-1}u - M^{-1}r. \end{aligned} \tag{3.19}$$

(3.20)

α can be determined from the fact that $\tilde{t} \perp u$, which leads to

$$\alpha = \frac{u^H M^{-1} r}{u^H M^{-1} u}. \tag{3.21}$$

By taking different values for α and M some already well known methods are obtained.

1. If $M = I$, then $\alpha = 0$ and $\tilde{t} = -r$. This is simply Arnoldi's method. $M = I$ can also be viewed as solving the correction equation with a 1-step Krylov-Method.

2. If $\alpha = 0$, then a Davidson method (cf. [28]) with preconditioner M is obtained. For $M = D - \theta I$, where D is the diagonal of A, it is the original Davidson method.

3. By choosing α as in equation (3.21) an instance of Jacobi-Davidson is obtained.

4. If $M = A - \sigma I$, then Jacobi-Davidson is similar to inverse iteration with Rayleigh shifts. This can be seen in the following way. With $M = A - \sigma I$ equation (3.19) becomes

   $$\tilde{t} = \alpha(A - \sigma I)^{-1}u - u$$

   Remember that $r = Au - \sigma u$. Since t is orthogonalized to the subspace V which contains u, it is equivalent to obtain $\tilde{t}$ from

   $$(A - \sigma I)\tilde{t} = u$$

   which is just the directon of inverse iteration. Since t is used to extend the subspace V_k, the direction of inverse iteration is contained in V_{k+1}. Inverse iteration converges cubically with Rayleigh shifts. Therefore we can expect asymptotically cubic convergence for Jacobi-Davidson if the correction equation is solved exactly.

The last fact is a very interesting property of the correction equation. If equation (3.18) is solved exactly we can expect cubic convergence if A is Hermitian. But if we only apply some steps of an iterative solver to equation (3.18) we may still obtain a good convergence behaviour if the Krylov solver steers $\tilde{t}$ into the direction of the exact solution t. In fact numerical experiments show that only some steps of a Krylov solver applied to the correction equation are sufficient to obtain a good convergence behaviour of Jacobi-Davidson.

Chapter 4

Nonlinear eigenvalue problems

4.1 Introduction

From now on we want to discuss arbitrary nonlinear eigenvalue problems of the type

$$T(\lambda)x = 0.$$

Definition 4.1 (Nonlinear Eigenvalue Problem)
Let $T(\lambda) \in \mathbb{C}^{n\times n}, \lambda \in \mathbb{C}$ be family of complex matrices. If the pair (λ, x) is a nontrivial solution of the equation

$$T(\lambda)x = 0, \tag{4.1}$$

such that $x \neq 0$, the value λ is called an eigenvalue of $T(\cdot)$ and x is called a corresponding right eigenvector. A left eigenvector y corresponding to an eigenvalue λ fulfills the equation

$$y^H T(\lambda) = 0.$$

If nothing else is stated, we are speaking of right eigenvectors throughout this context.

We will denote the multiplicity of an eigenvalue λ of $T(\cdot)$ according to the algebraic multiplicity of the zero-eigenvalue of $T(\lambda)x = \mu x$. If λ is an eigenvalue of $T(\cdot)$ there is at least one zero-eigenvalue $\mu = 0$ since $\det T(\lambda) = 0$.

Problems of type (4.1) arise in many applications, for example in bifurcation problems, damped vibration analysis or fluid-structure interaction. The linear eigenvalue problem $Ax = \lambda Bx$ is just a special case of (4.1) with

$$T(\lambda) = \lambda B - A.$$

Another common example resulting from damped vibration analysis is the quadratic eigenvalue problem where

$$T(\lambda) = \lambda^2 M + \lambda C + K.$$

Later we will also deal with rational problems of the form

$$T(\lambda) = -K + \lambda M + \sum_{j=1}^{K} \frac{\rho_0 \lambda}{k_j - \lambda m_j} C.$$

An important case occurs if $T(\lambda)$ is a family of real symmetric matrices where λ is chosen from a suitable subset $J \subset \mathbb{R}$. We will call this case a symmetric eigenvalue problem. For symmetric eigenvalue problems there exist minmax results which are very similar to the minmax results given in chapter 2 for linear Hermitian eigenvalue problems. These results will show a strong relationship between symmetric nonlinear and symmetric linear eigenvalue problems.

In order to reveal some of the properties of symmetric nonlinear eigenvalue problems we will first analyze quadratic overdamped problems. For these kind of problems Duffin (cf. [6]) could show minmax theorems similar to those for linear eigenvalue problems. These were later extended by Rogers to finite-dimensional overdamped problems (cf. [18]), and by Rogers, Langer, Turner, Hadeler and Werner (cf. [8],[9],[19],[39],[34], [35],[14]) to infinite-dimensional overdamped problems. Extensions to symmetric infinite-dimensional nonoverdamped problems were performed by Voss and Werner in [37] and Voss [36]. Overdamping will be discussed later in this context.

4.2 Analysis of quadratic problems

Let us consider a damped vibrating system without external forces. It can be described with the following differential equation:

$$A\ddot{u} + B\dot{u} + Cu = 0$$

C is the stiffness matrix and A is the mass matrix of the system. The matrix B describes the damping of the system. For the analysis performed in this section we assume that A, B and C are real symmetric positive semidefinite. With the ansatz $u = xe^{\lambda t}$ we finally arrive at the equation

$$T(\lambda)x = 0$$

with

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

For the following analysis we will need the Rayleigh quotients of the matrices A, B and C. Let $v \in \mathbb{R}^n \backslash \{0\}$. Then

$$a(v) := \frac{\langle Av, v\rangle}{\langle v, v\rangle}, \qquad b(v) := \frac{\langle Bv, v\rangle}{\langle v, v\rangle}, \qquad c(v) := \frac{\langle Cv, v\rangle}{\langle v, v\rangle}$$

In [6] Duffin investigated the eigenvalue distribution of $T(\cdot)$ with respect to the so called overdamping property

$$b^2(v) - 4a(v)c(v) > 0 \quad \forall v \in \mathbb{R}^n \backslash \{0\}. \tag{4.2}$$

In the case of a one dimensional system the overdamping property states that the two roots of the equation $ax^2 + bx + c = 0$ with real positive coefficients a, b and c lie on the negative real axis. Duffin showed that the overdamping property (4.2) leads to similar properties for multi-dimensional systems.

For the characterization of the roots of $T(\cdot)$ we need the following definition.

Definition 4.2 (Primary Functional)
For each non-zero vector v the primary functional is defined as

$$p(v) = -\frac{2c(v)}{b(v) + d(v)} \tag{4.3}$$

where

$$d(v) = \sqrt{b^2(v) - 4a(v)c(v)}.$$

Hence, $p(v)$ is just the largest root of the equation

$$p^2(v)a(v) + p(v)b(v) + c(v) = 0.$$

and we have the following lemma.

Lemma 4.3
$p(v) = x$ if and only if x satisfies

$$x^2 a(v) + xb(v) + c(v) = 0 \tag{4.4}$$

and

$$2xa(v) + b(v) > 0 \tag{4.5}$$

If a solution (λ, v) of $T(\lambda)v = 0$ corresponds to the functional p such that $\lambda = p(v)$ Duffin called the eigenvalue λ a primary eigenvalue and the eigenvector v a primary eigenvector. It can be shown that the primary eigenvectors to different eigenvalues are independent.

Lemma 4.4
Let $u_1, u_2, \ldots, u_m$ be primary eigenvectors with corresponding eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_m$. Then

$$\lambda_1 < p(u_1 + u_2 + \cdots + u_m) < \lambda_m$$

and the vectors are independent.

Proof:
[6]

Hence, the functional p leads to similar results as the Rayleigh quotient in the linear case. So it is interesting to investigate whether there also exist minmax properties for p which describe the eigenvalues of $T(\cdot)$. Indeed, this will be possible.

For the further analysis we need the definition of a number $P(Y)$ related to each subspace Y of dimension greater than zero.

$$P(Y) := \sup_{v \in Y} p(v).$$

Then the i^{th} primary minmax value is defined as

$$k_i = \inf_{Y \in H_i} P(Y),$$

where H_i is the set of all i-dimensional subspaces of the Hilbertspace H which is in this section the $\mathbb{R}^n$. Since p is bounded the numbers k_i are finite. The inf and sup values are reached inside the corresponding sets. So it is possible to use min and max values and we have

$$k_i = \min_{Y \in H_i} \max_{v \in Y} p(v).$$

Now it is possible to show the relation between minmax values and primary eigenvalues of $T(\cdot)$. We will give here exactly Duffins proof as it appeared in [6]. It is based on the idea to construct a sequence of subspaces which converge against the subspace U, for which $P(U)$ is a minmax value. The limit process will reveal that the minmax value is identical to an eigenvalue. The proof heavily depends on the structure of quadratic overdamped eigenvalue problems. The proof of the minmax theorem for arbitrary symmetric nonoverdamped problems uses the correspondence of nonlinear eigenvalue problems with linear problems and will hence not depend on the special structure of the problem any more.

Lemma 4.5
A primary minmax value is a primary eigenvalue

Proof:
Let k denote the i^{th} minmax value, so $k = P(U)$ where U is i-dimensional. The transformation $J_m = I + m^{-1}T(k)$ is nonsingular if m is a sufficiently large positive integer. Thus $Y_m = J_m U$ is an i-dimensional subspace. Let $P(Y_m) = p(y_m)$ where y_m is in Y_m and $\|y_m\| = 1$. For some u_m in U then, $y_m = u_m + m^{-1}T(k)u_m$. It is clear from this relation that $\|u_m\|$ is bounded as $m \to \infty$. Thus there is a vector u in U such that $u_m \to u$ as $m \to \infty$ through a suitable sequence of integers. Consequently $y_m \to u$, and so $\|u\| = 1$.

Let v be a given vector. Then if $p = p(v), p^2a(v) + pb(v) + c = 0$, so if l is an arbitrary real number, $-(p-l)(pa(v) + la(v) + b(v)) = l^2a(v) + lb(v) + c(v) = \frac{\langle T(l)v,v\rangle}{\langle v,v\rangle}$. Let $v = w + z$ and let $l = p(w)$. Then

$$-(p-l)(pa(v) + la(v) + b(v)) = \frac{2\langle T(l)w, z\rangle + \langle T(l)z, z\rangle}{\langle v, v\rangle} \tag{4.6}$$

In (4.6) take $v = y_m$, $w = u_m$, and $z = m^{-1}T(k)u_m$. Then $p(y_m) = P(Y_m) \geq P(U) \geq p(u_m)$, so $p-l = p(y_m)-p(u_m) \geq 0$. Also $pa(y_m)+la(y_m)+b(y_m) \to 2p(u)a(u)+b(u) > 0$. Thus for m sufficiently large the left side of (4.6) can not be positive.
Multiply (4.6) by m and substitue on the right, so

$$0 \geq \langle T(l)u_m, T(k)u_m\rangle + m^{-1}\langle T(l)T(k)u_m, T(k)u_m\rangle. \tag{4.7}$$

It was seen above that $p(y_m) \geq k \geq p(u_m)$. Since $y_m \to u$ and $u_m \to u$ it follows that $k = p(u)$. Hence $l = p(u_m) \to k$. Thus in the limit (4.7) becomes $0 \geq \langle T(k)u, T(k)u\rangle$. Therefore $T(k)u = 0$ and u is a primary eigenvector and k is a primary eigenvalue.

With this lemma it is easy to prove the following theorem.

THEOREM 4.6
There is an independent set of n primary eigenvectors $u_1, u_2, \ldots, u_n$. The corresponding eigenvalues are the minmax values $k_1, k_2, \ldots, k_n$. Any other primary eigenvector u is a linear combination of vectors of the set having the same eigenvalue as u.

Proof:
[6]

A direct conclusion is

COROLLARY 4.7
The ordered set of n primary minmax values and the ordered set of n primary eigenvalues are identical.

Proof:
[6]

Up to now we only discussed values resulting from the larger solution of $p^2a(v)+pb(v)+c(v) = 0$. Now we also want to consider the other solution of the quadratic equation. The secondary functional $s(v)$ is defined for a vector v if and only if $a(v) \neq 0$. The definition is

$$2sa(v) + b(v) = -d(v). \tag{4.8}$$

From the definition it follows directly that the secondary functional is a solution of $s^2a(v) + sb(v) + c(v) = 0$. Similar to primary eigenvalues and eigenvectors a vector w is

said to be a secondary eigenvector if $T(h)w = 0$ and $h = s(w)$. Then h is a secondary eigenvalue. To analyze secondary eigenvalues we consider the reciprocal problem

$$T(\tilde{\lambda})x = (A + \tilde{\lambda}B + \tilde{\lambda}^2 C)x = 0.$$

The primary functional for this problem is

$$p^0(v) = \frac{-2a(v)}{b(v) + d(v)}.$$

If $a(v) \neq 0$, then $s = 1/p^0$. Hence, the nonzero primary eigenvalues k^0 with eigenvectors w of the reciprocal problem correspond to the secondary eigenvalues and eigenvectors of the original problem by $h = 1/k^0$ which is clear because if $\tilde{\lambda} \neq 0$, the original problem is achieved by multiplying the reciprocal problem with $1/\tilde{\lambda}^2$ and setting $\lambda = 1/\tilde{\lambda}$. This back transformation of the reciprocal problem is not possible for zero eigenvalues. A zero eigenvalue of the reciprocal problem leads to

$$Aw = 0.$$

Hence, if the rank of A is $r < n$, then the last equation has nontrivial solutions and the reciprocal problem has a primary eigenvalue 0 whose eigenvectors form a $n - r$-dimensional subspace. This leads to the following theorem

THEOREM 4.8
Let r be the rank of A. Then there is an independent set of r vectors $w_1, w_2, \ldots, w_r$. Each vector of the set is a secondary eigenvector. Any other secondary eigenvector is a linear combination of vectors of the set with the same eigenvalue.

A further important result is the following theorem.

THEOREM 4.9
The range of the primary functional and the range of the secondary functional have no common values.

Proof:
[6]

The roots of $\det(T(\lambda))$ do not change by congruence transformations. It is always possible to transform $\lambda^2 A + \lambda B + C$ such that $\tilde{A} = H^H A H$ is diagonal with the $n - r$ zero eigenvalues lying on the first $n-r$ diagonal elements of $\tilde{A}$ and $\tilde{B} = H^H B H$ having a diagonal $n - r \times n - r$ leading principial submatrix. The diagonal elements of this leading principal submatrix of $\tilde{B}$ can not be zero because otherwise the overdamping condition (4.2) is not fulfilled any more. Hence, the polynomial $\det(T(\lambda))$ is of degree $n + r$.

Together with the following lemma we will be able to characterize all eigenvalues of $T(\cdot)$.

LEMMA 4.10
For a given constant k, let R be the rank of the matrix $k^2A + kB + C$. *Then* $\det(T(\lambda))$ *has a zero of multiplicity* $n - R$ *at* $\lambda = k$.

Proof:
[6]

Hence, we have a direct correspondence between the multiplicity of the roots of $\det(T(\lambda))$ and the multiplicity of eigenvalues of $T(\cdot)$. The primary eigenvalues describe n eigenvalues of $T(\cdot)$, and the secondary eigenvalues describe r eigenvalues of $T(\cdot)$. In addition primary and secondary eigenvalues are distinct. Together with the last lemma we can conclude

THEOREM 4.11
A necessary and sufficient condition that k be a zero of $\det(T(\lambda))$ *of multiplicity j is that k be an eigenvalue of multiplicity j. This accounts for* $n+r$ *real zeros of* $\det(T(\lambda))$*; there are no other zeros.*

From the analysis the following last lemma can easily be concluded which shows that the eigenvalue distribution for multi-dimensional quadratic overdamped systems is similar to one-dimensinal overdamped systems.

LEMMA 4.12
The eigenvalues of $T(\cdot)$ *are negative or zero valued. Every primary eigenvalue exceeds every secondary eigenvalue.*

Proof:
[6]

There are two important aspects in Duffins analysis which need to be further investigated for a generalization of his results. The first one is the functional p. We will show later that this is a generalization of the Rayleigh quotient for linear systems and give a proper definition for arbitrary symmetric nonlinear eigenvalue problems. The second important aspect is the overdamping condition on which his proofs depend. The generalized overdamping condition has a strong influence on how eigenvalues will be counted.

4.3 The Rayleigh functional

In this section we want to introduce a generalization of the Rayleigh quotient for linear systems. We do not restrict the definition to finite dimensional Hilbert spaces because

the minmax theory presented later will also be valid for infinite dimensional Hilbert spaces.

Let $T(\lambda)$ be a selfadjoint and bounded operator on a real Hilbert space H for every λ in an open interval J. Then we define the function f as

$$f : \begin{cases} J \times H & \to & \mathbb{R} \\ (\lambda, x) & \mapsto & \langle T(\lambda)x, x\rangle \end{cases} \tag{4.9}$$

We assume that f is continously differentiable, and that for every fixed $x \in H^0$, $H^0 := H\backslash\{0\}$, the real equation

$$f(\lambda, x) = 0$$

has at most one solution $\lambda =: p(x)$ in J. Then equation (4.9) implicitly defines a functional p on a proper subset D of H^0 which we call the Rayleigh Functional. We further assume that

$$\frac{\partial}{\partial\lambda} f(p(x), x) > 0 \quad \forall x \in D. \tag{4.10}$$

This condition is always fulfilled if $T'(\lambda)$ is positive definite. If (λ, u) is an eigenpair of $T(\cdot)$ and $\lambda \in J$ we have $\lambda = p(u)$, since there is at most one solution of $f(\lambda, x) = 0$ in J. The derivate of $p(x)$ is given as

$$\nabla p(x) = \frac{-2T(p(x))x}{\langle T'(p(x))x, x\rangle}.$$

Hence $\nabla p(u) = 0$ if $u \in D$ is an eigenvector of $T(\cdot)$.

This definition of the Rayleigh functional is a true generalization of the Rayleigh quotient. So consider the linear eigenvalue problem $Ax = \lambda Bx$, where A is symmetric and B is symmetric positive definite. With $T(\lambda) = \lambda B - A$ we have

$$f(\lambda, x) = \langle(\lambda B - A)x, x\rangle.$$

The solution of $f(\lambda, x) = 0$ is just the usual Rayleigh quotient

$$p(x) = \frac{\langle x, Ax\rangle}{\langle x, Bx\rangle}.$$

and

$$\frac{\partial}{\partial\lambda} f(p(x), x) = \langle x, Bx\rangle > 0 \quad \forall x \in H^0.$$

Another example is the primary functional (4.3) introduced by Duffin for the quadratic overdamped eigenvalue problem

$$f(\lambda, x) = \langle(\lambda^2 A + \lambda B + C)x, x\rangle = 0.$$

The primary functional was defined such that

$$\frac{\partial}{\partial\lambda} f(p(x), x) = 2pa(x) + b(x) > 0$$

Hence, the primary functional is a Rayleigh functional. The secondary functional defined from (4.8) does not fulfill condition (4.10), since

$$\frac{\partial}{\partial\lambda} f(s(x), x) = 2sa(x) + b(x) = -d(x) < 0.$$

But this problem can easily be overcome by using $-T(\lambda)$ instead of $T(\lambda)$.

4.4 Overdamping

Duffin's analysis was based on the overdamping condition (4.2) for quadratic eigenvalue problems. In this section we want to motivate a general definition of overdamping and want to show why the minmax theory introduced by Duffin for quadratic problems and extended by Rogers for arbitrary symmetric finite-dimensional problems must fail for nonoverdamped problems

The overdamping condition for the quadratic eigenvalue problem

$$T(\lambda)x = \lambda^2 Ax + \lambda Bx + Cx = 0$$

is defined as

$$b^2(x) - 4a(x)c(x) > 0, \quad \forall x \in \mathbb{R}^n.$$

In this case all eigenvalues lie on the real axis as Duffin's analysis showed. Now we consider the example

$$T(\lambda) = \lambda^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \lambda \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}.$$

The eigenvalues of $T(\cdot)$ are $-1+i$, $-1-i$, -1, -4. Since there are two complex conjugate eigenvalues the problem can not be overdamped and in fact in the direction of e_2 we have.

$$b^2(e_2) - 4a(e_2)c(e_2) = -4 < 0.$$

Hence, the Rayleigh functional has no real solutions in the direction of e_2. So the domain of the Rayleigh functional must be restricted to the directions v in which the problem is overdamped and therefore admits real solutions of the Rayleigh functional. As conclusion in order to have a chance of a minmax characterization for the nonoverdamped quadratic eigenvalue problem, the domain of the Rayleigh functional must be restricted to areas where the overdamping condition holds. For the case that $H = \mathbb{R}^n$

Barston first investigated minmax properties of nonoverdamped quadratic eigenvalue problems (cf. [2]).

To generalize the results of the quadratic problem we want to call a symmetric eigenvalue problem overdamped if for the domain D of the Rayleigh functional we have $D = H^0$. For overdamped problems Rogers extended Duffins analysis for symmetric eigenvalue problems in $\mathbb{R}^n$. He proved following theorem for the minmax values

$$k_i = \min_{V \in H_i} \max_{x \in V \setminus \{0\}} p(x)$$

THEOREM 4.13
The value k is a minmax value if, and only if, k is in the range of p and (k, v) is an eigenpair for some non-zero v.

Proof:
[18]

For nonoverdamped problems the last theorem must fail as the following example shows.

Consider following eigenvalue problem

$$Ax = \lambda x$$

with

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}, \lambda \in J = (1.5, 3.5) \subset \mathbb{R}.$$

Since λ is restricted to the open interval $(1.5, 3.5)$ the problem is not linear. The Rayleigh functional

$$p(x) = \frac{x^H A x}{x^H x}$$

is defined on the open set $D \in \mathbb{R}^5 \setminus \{0\}$ of vectors who are mapped into J by p. Because $D \neq \mathbb{R}^n \setminus \{0\}$ it is a nonoverdamped problem. The smallest eigenvalue of A in J is 2. But

$$\inf_{x \in D} p(x) = 1.5,$$

which is not even contained in J. Hence, the minmax theory of Rogers can not directly be extended to nonoverdamped problems. The solution of Voss and Werner was to introduce a new numbering of the eigenvalues of $T(\cdot)$. In Rogers theory the smallest eigenvalue $\lambda_1 \in J$ of $T(\cdot)$ corresponds to the first minmax value k_1. As we have just seen, this is not applicable in this example of a nonoverdamped problem.

Voss and Werner oriented the numbering of the eigenvalues of $T(\cdot)$ according to the corresponding linear problem

$$T(\lambda)x = \mu x. \tag{4.11}$$

If λ is an eigenvalue of $T(\cdot)$ and $T(\lambda) + \nu(\lambda)I$ is completely continuous for a function $\nu(\lambda)$ there exists a zero eigenvalue $\mu = 0$ of (4.11) which can be characterized by linear minmax theory as

$$0 = \mu_i = \sup_{V \in H_i} \min_{x \in V \setminus \{0\}} \frac{\langle x, T(\lambda)x \rangle}{\langle x, x \rangle},$$

Then the eigenvalue λ of $T(\cdot)$ receives the number i.

4.5 A general minmax theory

We are now ready to introduce a minmax theory for nonoverdamped eigenvalue problems. The results given in this chapter were developed by Voss and Werner in [37].

We consider the nonlinear eigenvalue problem

$$T(\lambda)x = 0,$$

where for every λ in a real open interval J the operator $T(\lambda)$ is selfadjoint and bounded. Further let the function $f(\lambda, x)$ and the corresponding Rayleigh Functional be defined as in chapter 4.3 with the property (4.10).

Let H_j denote the set of all j-dimensional subspaces of H and

$$\begin{aligned} H_{j,D} &:= \{V \in H_j : V \cap D \neq \varnothing\} \\ V^1 &:= \{v \in V : \|v\| = 1\} \\ \mu_V(\lambda) &:= \min_{v \in V^1} \langle T(\lambda)v, v \rangle, \\ \mu_j(\lambda) &:= \sup_{V \in H_j} \mu_V(\lambda) \\ \lambda(V) &:= \sup_{u \in V \cap D} p(u) \end{aligned}$$

We further assume that for every fixed $\lambda \in J$ there exists $\nu(\lambda) > 0$ such that the linear operator $T(\lambda) + \nu(\lambda)I$ is completely continuous. Then the essential spectrum of $T(\lambda)$ contains only the point $-\nu(\lambda)$ and the following conditions hold:

(A1) If $\mu_n = 0$ for some $n \in \mathbb{N}$ and $\lambda \in J$, then $\mu_j(\lambda) = \max_{V \in H_j} \mu_V(\lambda)$ $(j = 1, \ldots, n)$ and $\mu_1 \geq \mu_2 \geq ... \geq \mu_n$ are the first n eigenvalues of $T(\lambda)$.

(A2) If $\mu = 0$ is an eigenvalue of $T(\lambda)$, then there exists $n \in \mathbb{N}$ with $\mu_n(\lambda) = 0$.

As stated in chapter 4.2 the eigenvalues of $T(\cdot)$ shall be enumerated according to the eigenvalues of the corresponding linear problem $T(\lambda)x = \mu x$. Hence, we have following definition.

DEFINITION 4.14 (N-TH EIGENVALUE)
If $\lambda \in J$ is an eigenvalue of $T(\cdot)$, then by (A2) there exists $n \in \mathbb{N}$ such that $\mu_n(\lambda) = 0$. In this case we call λ an n-th eigenvalue of $T(\cdot)$. If conversely $\mu_n(\lambda) = 0$ for some $n \in \mathbb{N}$ and $\lambda \in J$ then (A1) implies that λ is an n-th eigenvalue of $T(\cdot)$.

With this notation we can introduce the main theorem of this section.

THEOREM 4.15

(i) For every $n \in \mathbb{N}$ there is at most one n-th eigenvalue $\lambda_n \in J$ of $T(\cdot)$ which can be characterized by

$$\lambda_n = \min_{V \in H_{n,D}} \lambda(V) \tag{4.12}$$

The set of eigenvalues of $T(\cdot)$ is at most countable.

(ii) If there exist the m-th and n-th eigenvalue λ_m and λ_n in $J (m < n)$, then J containts the k-th eigenvalue $\lambda_k (m < k < n)$ too, and

$$\inf J < \lambda_m \leq \lambda_{m+1} \leq \cdots \leq \lambda_n < \sup J.$$

(iii) Let λ_n be defined by

$$\lambda_n := \inf_{V \in H_{n,D}} \sup_{v \in V \cap D} p(v).$$

If $\lambda_n \in J$ for some $n \in \mathbb{N}$, then λ_n is the n-th eigenvalue of $T(\cdot)$ and (i) holds.

Here we will only give a sketch of the proof. The complete proof can be found in [37]. We need the following lemmas for the proof.

LEMMA 4.16
For every $\lambda \in J$ and $u \in D$ we have

$$\lambda \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} p(u) \Leftrightarrow \langle T(\lambda)u, u \rangle \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} 0. \tag{4.13}$$

Proof:
This lemma follows directly from the fact that $f(\lambda, u)$ has at most one solution in J and that $\frac{\partial}{\partial \lambda} f(p(u), u) > 0$ for $u \in D$.

LEMMA 4.17
Let $\mu_V(\lambda) = 0$ for some $\lambda \in J$ and $V \in H_n$. Then $V \in H_{n,D}$ and

$$\lambda = \lambda(V) = \max_{u \in V \cap D} p(u) \tag{4.14}$$

Proof:
There exists $\bar{u} \in V$ such that $\langle T(\lambda)u, u\rangle \geq \langle T(\lambda)\bar{u}, \bar{u} = 0$ for every $u \in V$. Hence $\bar{u} \in D \cap V$ and $V \in H_{n,D}$. (4.14) follows from (4.13).

The next lemma gives the correspondence between $\lambda(V)$ and $\mu_V(\lambda)$, hence the connection between the nonlinear eigenvalue problem $T(\lambda)x = 0$ and the linear problem $T(\lambda)x = \mu x$.

Lemma 4.18
For every $\lambda \in J$ and $V \in H_{n,D}$ we have

$$\lambda \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} \lambda(V) \Leftrightarrow \mu_V(\lambda) \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} 0. \tag{4.15}$$

Proof:
[37]

Now we are ready to proof (i) of Theorem 4.15. Let λ be an n-th eigenvalue of $T(\cdot)$. Then $\mu_n(\lambda) = 0$ and (A1) implies the existence of some $\bar{V} \in H_n$ such that

$$\mu_V(\lambda) \leq \mu_{\bar{V}}(\lambda)$$

for every $V \in H_n$. From Lemma 4.17 we know that $\bar{V} \in H_{n,D}$ and

$$\lambda = \lambda(\bar{V}) = \max_{u \in \bar{V} \cap D} p(u).$$

From Lemma 4.18 it follows that $\lambda(V) \geq \lambda$ for all $V \in H_{n,D}$ which concludes the proof of (i). It is remarkable that the space $\bar{V}$ for which the minimum is achieved is just the space spanned by the eigenvectors to the n largest eigenvalues of $T(\lambda_n)$. This shows the strong connection between the nonlinear eigenvalue problem and the corresponding linear problem. We will exploit this fact later when we present solution methods for nonlinear eigenvalue problems. For the proof of (ii) we first observe that $\lambda_m \leq \lambda_n$ which follows directly from the minmax characterization (i). Analogue to an argumentation given in [7] we have

$$\mu_m(\lambda_n) \geq \mu_{m+1}(\lambda_n) \geq \cdots \geq \mu_n(\lambda_n) = 0 = \mu_m(\lambda_m) \geq \mu_{m+1}(\lambda_m) \geq \cdots \geq \mu_n(\lambda_m)$$

From the continuity of μ_k in J it follows that every of the equations $\mu_k(\lambda) = 0, k = m+1, \ldots, n-1$ has a root λ_k^* in J with $\mu_k(\lambda_k^*) = 0$, which concludes part (ii) of Theorem 4.15. The proof of (iii) can be found in ([37]).

We conclude this section with a theorem answering the question whether for some $V \in H_{n,D}$

$$V \in \bar{H}_n := \{W \in H_n : W^0 \subset D\}$$

and whether $\bar{V} \in \bar{H}_n$ if $\lambda_n = \lambda(\bar{V})$.

THEOREM 4.19
Let λ_j be defined as in (iii) of Theorem 4.15 and $\lambda_1, \lambda_n \in J$ for some $n \in \mathbb{N}$. Then there are n eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ of $T(\cdot)$ in J. Each λ_j is characterized by (i) of Theorem 4.15 and each $V \in H_{j,D}$ with $\lambda_j = \lambda(V)$ belongs to $\bar{H}_j (j = 1, \ldots, n)$. One can write

$$\lambda_j = \min_{V \in \bar{H}_j} \max_{u \in V^1} p(u), j = 1, \ldots, n$$

Proof:
[37]

Chapter 5

Methods for nonlinear problems

In this chapter we will discuss solution methods for the nonlinear eigenvalue problem (4.1). In the first part we will review tradtitional methods which require several factorizations of $T(\lambda)$ in every step to find an eigenvalue approximation. An exception is the residual inverse iteration introduced by Neumaier (cf. [16]). It needs only one factorization of $T(\lambda)$. But then only linear convergence can be proved. A not so famous but very efficient method is the guarded iteration introduced by Werner (cf. [38]). It is based on the minmax theory for symmetric nonlinear eigenvalue problems and allows to iterate for the k-th eigenvalue of $T(\cdot)$ while the other methods presented for small dense problems only allow to iterate for an eigenvalue near a given start approximation independent of its number. Then we will present a first approach for large sparse nonlinear eigenvalue problems which was first introduced by Ruhe (cf. [24]) and then worked out in detail by Hager, Ruhe and Wiberg in ([10],[11],[27]). But this method seems to be only suitable for slightly nonlinear problems in which the nonlinearity has no strong influence. In the last part of this chapter we will introduce a new method for symmetric nonlinear eigenvalue problems based on a Jacobi-Davidson type approach (cf. [3]). It overcomes the necessity to perform matrix factorizations, but still delivers a fast convergence. A good working restart strategy is also demonstrated.

5.1 Methods for dense problems

The methods discussed now are suitable for small dense nonlinear eigenvalue problems. They are based on a sequence of factorizations of $T(\lambda)$ or a sequence of successive linear eigenvalue problems to find an eigenvalue of the nonlinear problem $T(\lambda)x = 0$. A good overview can also be found in [21].

5.1.1 Inverse iteration

Inverse iteration for nonlinear eigenvalue problems is a simple consequence of the application of Newtons's method to the equation

$$F\begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} T(\lambda)x \\ v^H x - 1 \end{pmatrix} \tag{5.1}$$

The ansatz

$$F\begin{pmatrix} x_k \\ \lambda_k \end{pmatrix} + F'\begin{pmatrix} x_k \\ \lambda_k \end{pmatrix}\begin{pmatrix} x_{k+1} - x_k \\ \lambda_{k+1} - \lambda_k \end{pmatrix} = 0$$

leads to

$$\begin{bmatrix} T(\lambda_k) & T'(\lambda_k)x_k \\ v_k^H & 0 \end{bmatrix}\begin{pmatrix} x_{k+1} - x_k \\ \lambda_{k+1} - \lambda_k \end{pmatrix} = -\begin{pmatrix} T(\lambda_k)x_k \\ v_k^H x_k - 1 \end{pmatrix}$$

where v_k is a suitable normalization vector in step k. If x_k is already normalized with respect to v_k the second row of the last equation yields $v_k^H x_{k+1} = v_k^H x_k$. From the first row we have

$$x_{k+1} = -(\lambda_{k+1} - \lambda_k)T(\lambda_k)^{-1}T'(\lambda)x_k.$$

By setting $u_{k+1} := T(\lambda_k)^{-1}T'(\lambda_k)x_k$, multiplying the last equation with v_k^H and using that $v_k^H x_{k+1} = v_k^H x_k$ we obtain.

$$\lambda_{k+1} = \lambda_k - \frac{v_k^H x_k}{v_k^H u_{k+1}}. \tag{5.2}$$

Algorithm 6 Inverse Iteration for Nonlinear Eigenvalue Problems

1: Start with v_1, x_1, λ_1
2: **for** $k = 1, \ldots$ until convergence **do**
3: $u_{k+1} = T(\lambda_k)^{-1}T'(\lambda_k)x_k$
4:
$$\lambda_{k+1} = \lambda_k - \frac{v_k^H x_k}{v_k^H u_{k+1}}$$
5: Choose a new normalization vector v_{k+1}.
6:
$$x_{k+1} = \frac{u_{k+1}}{v_{k+1}^H u_{k+1}}$$
7: **end for**

THEOREM 5.1
Let λ be a simple eigenvalue of $T(\cdot)$ with corresponding eigenvector x, $v_k = v$ for all $k \in \mathbb{N}$ and $v^H x_0 = v^H x = 1$. Then Algorithm 6 converges locally quadratic against the eigenpair (λ, x) if $T'(\lambda)$ is regular and 0 is an algebraic simple eigenvalue of $T'(\lambda)^{-1}T(\lambda)$.

Proof:
Since Algorithm 6 is derived from Newton's method it is sufficient to show that $F'(x,\lambda)$ is not singular. Let

$$F'(x,\lambda) = \begin{bmatrix} T(\lambda_k) & T'(\lambda_k)x_k \\ v^H & 0 \end{bmatrix} \begin{pmatrix} t \\ \mu \end{pmatrix} = 0. \tag{5.3}$$

We have to show that the only solution of (5.3) is $t = 0, \mu = 0$.

$\mu = 0$: From equation (5.3) we have

$$T(\lambda)t = 0.$$

and therefore $t = \alpha x, \alpha \in \mathbb{C}^n$. From $v^H t = 0$ it follows that $0 = v^H t = \alpha v^H x = \alpha$. Hence, we have $t = 0$.

$\mu \neq 0$: In this case the first row of equation (5.3) says that

$$T(\lambda)t = -\mu T'(\lambda)x.$$

By premultiplying this equation with $T'(\lambda)^{-1}$ we arrive at

$$T'(\lambda)^{-1}T(\lambda)t = -\mu x \neq 0.$$

Since $T(\lambda)x = 0$ it follows that

$$(T'(\lambda)^{-1}T(\lambda))^2 t = -\mu(T'(\lambda)^{-1}T(\lambda))x = 0$$

But t can not be a principle eigenvector of order 2 of $T'(\lambda)^{-1}T(\lambda)$ corresponding to 0, because 0 is a simple eigenvalue of $T'(\lambda)^{-1}T(\lambda)$.

In [21] Ruhe proposes to use as v_k the vectors $v_k = T(\lambda_k)^H y_k$, where y_k is an approximation for a left eigenvector. Then the update (5.2) becomes

$$\lambda_{k+1} = \lambda_k - \frac{y_k^H T(\lambda_k)x_k}{y_k^H T'(\lambda_k)x_k}.$$

This is the Rayleigh functional proposed by Lancaster (cf. [21]) for nonlinear eigenvalue problems. The last equation can be seen as one Newton step to solve the equation

$$\tilde{f}_k(\lambda) := y_k^H T(\lambda)x_k = 0.$$

For symmetric matrices it is also possible to update the eigenvalue approximation in every step with the Rayleigh Functional. Hence, we use

$$\lambda_{k+1} = p(u_{k+1}),$$

where p is defined as in chapter 4.3. If $T(\cdot)$ is linear we arrive at inverse iteration for linear eigenvalue problems which converges cubically. This cubic convergence can also be shown for symmetric nonlinear eigenvalue problems if the Rayleigh functional update is used (cf. [20]). Hence, the inverse iteration derived here is just an extension of the already known inverse iteration for linear eigenvalue problems. The problem of inverse teration for nonlinear problems is the necessity to perform a factorization of $T(\lambda)$ in every step to solve the system $u_{k+1} = T(\lambda_k)^{-1}T'(\lambda_k)x_k$. In the linear case it is possible to work with a fixed shift σ. But if $T(\cdot)$ is not linear the iteration $u_{k+1} = T(\sigma)^{-1}T'(\lambda_k)x_k$ leads to convergence towards an eigenvector of the linear problem

$$T(\sigma)x = \mu T'(\lambda^*)x,$$

where λ^* is the limit of the sequence $\{\lambda_k\}$. Hence, using a fixed shift leads to a wrong convergence behaviour for nonlinear eigenvalue problems. In [16] Neumaier proposed a possibility to overcome this problem by performing a residual inverse iteration, which he derived in the following way.

We now assume that $T(\lambda)$ is twice continuous differentiable. If we choose the same vector $v_k = v$ in every step of Algorithm 6, the iteration can be written as

$$\begin{aligned} u_{k+1} &= T(\lambda_k)^{-1}T'(\lambda_k)x_k && (5.4)\\ x_{k+1} &= \frac{u_{k+1}}{v^H u_{k+1}} && (5.5)\\ \lambda_{k+1} &= \lambda_k - \frac{1}{v^H u_{k+1}} && (5.6) \end{aligned}$$

We want to calculate the update dx such that $x_{k+1} = x_k - dx$.

$$\begin{aligned} dx &= x_k - x_{k+1}\\ &= x_k + (\lambda_{k+1} - \lambda_k)u_{k+1}\\ &= x_k + (\lambda_{k+1} - \lambda_k)T(\lambda_k)^{-1}T'(\lambda_k)x_k\\ &= T(\lambda_k)^{-1}(T(\lambda_k) + (\lambda_{k+1} - \lambda_k)T'(\lambda_k))x_k\\ &= T(\lambda_k)^{-1}T(\lambda_{k+1})x_k + O((\lambda_{k+1} - \lambda_k)^2) \end{aligned}$$

By neglecting the second order term we arrive at the update step

$$x_{k+1} = x_k - T(\lambda_k)^{-1}T(\lambda_{k+1})x_k.$$

The great advantage of Neumaiers approach is that a fixed shift σ can be used in this iteration without loosing convergence to an eigenpair of $T(\cdot)$. With a fixed shift the update becomes

$$x_{k+1} = x_k - T(\sigma)^{-1}T(\lambda_{k+1})x_k.$$

The value λ_{k+1} has to be determined before the update can be performed. Neumaier proposed two possibilities. If $T(\lambda)$ is Hermitian and λ real we can use the update

$$\lambda_{k+1} = p(x_k). \tag{5.7}$$

For arbitrary matrices $T(\lambda)$ the solution λ_{k+1} of the real equation

$$v^H T(\sigma)^{-1} T(\lambda_{k+1}) x_k = 0, \tag{5.8}$$

which is closest to λ_k is used. To understand the update (5.8) we can see $v^H T(\sigma)^{-1}$ as approximation y_k for a left eigenvector y. Applying one step of Newtons's method to solve for

$$y_k^H T(\lambda_{k+1}) x_k = 0$$

results in

$$\lambda_{k+1} = \lambda_k - \frac{y_k^H T(\lambda_k) x_k}{y_k^H T'(\lambda_k) x_k},$$

which is again the Rayleigh functional proposed by Lancaster.

The speed of convergence of the residual inverse iteration is stated in the following theorem.

THEOREM 5.2
Let $T(\lambda)$ be twice continuous differentiable, $\hat{\lambda}$ be a simple zero of $\det T(\lambda) = 0$, and suppose that there is a corresponding eigenvector $\hat{x}$ normalized such that $v^H \hat{x} = 1$. Then the Residual Inverse Iteration converges for all σ sufficiently close to $\hat{\lambda}$, and we have

$$\frac{\|x_{k+1} - \hat{x}\|}{\|x_k - \hat{x}\|} = O(\sigma - \hat{\lambda}), \quad \|\lambda_{k+1} - \hat{\lambda}\| = O(\|x_k - \hat{x}\|^t),$$

where $t = 2$ if $T(\lambda)$ is Hermitian, $\hat{\lambda}$ is real and (5.7) is used. Otherwise if (5.8) is used, we have $t = 1$.

Proof:
[16]

5.1.2 Successive linear approximations

Another approch to solve (4.1) is to use linear approximations of $T(\lambda)$. If we have an approximation λ_k for an eigenvalue, we are seeking for a correction h and a corresponding vector x such that

$$T(\lambda_k - h)x = 0.$$

To approximate the solution, we linearize the last equation and obtain the generalized eigenvalue problem

$$T(\lambda_k)x = hT'(\lambda_k)x. \tag{5.9}$$

The eigenvalue h of smallest modulus is used as correction. Near an eigenvalue $\hat{\lambda}$ of $T(\cdot)$ this method is very similar to inverse iteration. This can be seen by rewriting equation (5.9) as

$$\frac{1}{h}x = T(\lambda_k)^{-1} T'(\lambda_k)x. \tag{5.10}$$

We are seeking for the eigenvalue of $T(\lambda_k)^{-1}T'(\lambda_k)$ with largest modulus. One step of inverse iteration is therefore similar to applying one step of a power method to find this eigenvalue.

5.1.3 QR type methods

Now we want to investigate a different approach which was introduced by Kublanovskaya (cf. [13]). It is based on the idea to find a root of $\det T(\lambda)$ by working on the QR decomposition of $T(\lambda)$.

Let

$$T(\lambda)P(\lambda) = Q(\lambda)R(\lambda)$$

be the QR-decomposition of $T(\lambda)P(\lambda)$ where $P(\lambda)$ is a permuation matrix which is chosen such that the diagonal elements of $R(\lambda)$ are ordered descending in magnitude. Hence, $|r_{11}| \geq \cdots \geq |r_{nn}|$. Then

$$|\det T(\lambda)| = |\det R(\lambda)|.$$

Therefore $\det T(\lambda) = 0$ is reached if $r_{nn} = 0$, since then we have $\det R(\lambda) = 0$. Let $f(\lambda)$ be defined as

$$f(\lambda) := r_{nn}(\lambda).$$

We want to derive a formulation of Newton's method to solve for $f(\lambda) = 0$. Let λ_k be the current approximation for an eigenvalue of $T(\cdot)$. If we differentiate

$$T(\lambda_k)P(\lambda_k) = Q(\lambda_k)R(\lambda_k)$$

according to λ we obtain

$$T'(\lambda_k)P(\lambda_k) + T(\lambda_k)P'(\lambda_k) = Q'(\lambda_k)R(\lambda_k) + Q(\lambda_k)R'(\lambda_k).$$

In a small neighborhood around λ_k the permutation matrix $P(\lambda_k)$ is constant. Therefore the derivate $P'(\lambda_k)$ is zero. By multiplying the last equation from the left with $Q^H(\lambda_k)$ and from the right with $R(\lambda_k)^{-1}$ we arrive at

$$Q^H(\lambda_k)T'(\lambda_k)P(\lambda_k)R^{-1}(\lambda_k) = Q^H(\lambda_k)Q'(\lambda_k) + R'(\lambda_k)R^{-1}(\lambda_k). \tag{5.11}$$

For the columns $q_j(\lambda)$ of $Q(\lambda)$ we have

$$\langle q_i(\lambda), q_j(\lambda)\rangle = \delta_{ij}.$$

By differentiating this equation it follows that

$$\langle q_i(\lambda), q_j'(\lambda)\rangle = -\overline{\langle q_j(\lambda), q_i'(\lambda)\rangle}.$$

Therefore $Q^H(\lambda_k)Q'(\lambda_k)$ is Skew-Hermitian with a zero diagonal. The last diagonal element of $R'(\lambda_k)R^{-1}(\lambda_k)$ equals $r'_{nn}(\lambda_k)/r_{nn}(\lambda_k)$. By considering that the diagonal of $Q^H(\lambda_k)Q'(\lambda_k)$ is zero and multiplying (5.11) from left and right with the n-th unity-vector e_n we obtain

$$e_n^H Q^H(\lambda_k)T'(\lambda_k)P(\lambda_k)R^{-1}(\lambda_k)e_n = \frac{r'_{nn}(\lambda_k)}{r_{nn}(\lambda_k)} \tag{5.12}$$

The Newton update to solve $f(\lambda) = 0$ is

$$\lambda_{k+1} = \lambda_k - \frac{r_{nn}(\lambda_k)}{r'_{nn}(\lambda_k)}.$$

Hence, we have

$$\lambda_{k+1} = \lambda_k - 1/e_n^H Q^H(\lambda_k)T'(\lambda_k)P(\lambda_k)R^{-1}(\lambda_k)e_n.$$

Approximations for left and right eigenvectors can be obtained as

$$y_k = Q(\lambda_k)e_n, \quad x_k = P(\lambda_k)R(\lambda_k)^{-1}e_n$$

5.1.4 Guarded iteration

The previous methods described for dense problems are suitable for symmetric and non-symmetric eigenvalue problems. Now we want to focus on a method especially for symmetric eigenvalue problems. Here we can apply results from the minmax theory introduced in chapter 4.5.

The guarded iteration was introduced by Werner in [38]. It is an iteration method for nonlinear symmetric eigenvalue problems that aims directly onto the k-th eigenvalue. Hence, it is very suitable for computing all eigenvalues in the range of the Rayleigh functional. For this section we use the notation introduced in 4.3 and assume that $T(\lambda)$ is of dimension $n < \infty$. We will derive a fixpoint iteration which converges locally cubically against the k^{th} eigenvalue of $T(\cdot)$ if $T'(\lambda)$ is positive definite. Otherwise we will still have local quadratic convergence.

Let $h_k(\lambda)$ be defined as

$$h_k : \begin{cases} J \supset J_k & \to & J \\ \lambda & \mapsto & p(x_k(\lambda)) \end{cases},$$

where x_k is an eigenvector corresponding to the k^{th} largest eigenvalue of $T(\lambda)$, hence $T(\lambda)x_k(\lambda) = \mu_k(\lambda)x_k(\lambda)$, where $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$. J_k is defined such that $x_k(\lambda) \in D$ for $\lambda \in J_k$. If $\mu_k(\lambda)$ is a multiple eigenvalue, the eigenvector $x_k(\lambda)$ can be chosen arbitrary in the eigenspace corresponding to μ_k.

If λ_k is the k^{th} eigenvalue of $T(\cdot)$ in J with the enumeration of eigenvalues given in chapter 4.5 we know that $T(\lambda_k)x_k(\lambda_k) = 0$ and therefore $p(x_k(\lambda_k)) = \lambda_k$ from the definition of the Rayleigh functional. Hence,

$$h_k(\lambda_k) = p(x_k(\lambda_k)) = \lambda_k$$

and λ_k is a fixpoint of $h_k(\lambda)$.

We will now investigate the properties of the fixpoint iteration

$$\alpha_{i+1} = h_k(\alpha_i) \tag{5.13}$$

First of all we must guarantee that $x_k(\alpha_k) \in D$ for all k. Otherwise the iteration is not well defined. This can be ensured by choosing the start value α_1 near the eigenvalue λ_k. Because $x_k(\lambda)$ is continuous near a simple eigenvalue λ_k (cf. Lemma 5.5), $x_k(\lambda_k) \in D$ and D is open the vector $x_k(\alpha_1)$ and all subsequent vectors $x_k(\alpha_i)$ will also be contained in D if λ_k is an attracting fixpoint of h_k.

We need the following lemma.

Lemma 5.3
Let $\lambda \in J$ and $k \in \mathbb{N}$. Then

$$\lambda \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} \lambda_k \Leftrightarrow \mu_k(\lambda) \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} 0. \tag{5.14}$$

Proof:
Since λ_k is a k-th eigenvalue in J it follows from the proof of Theorem 4.15 and Lemma 4.17 that there is a subspace $\bar{V} \in H_k$ such that

$$\lambda_k = \max_{x \in \bar{V} \cap D} p(x).$$

Hence, from $\lambda > \lambda_k$ it follows that $\lambda > p(x)$ for all $x \in \bar{V} \cap D$. This is equivalent to $\lambda > \lambda(\bar{V})$. From Lemma 4.18 it follows that $\mu_{\bar{V}}(\lambda) > 0$, which results in $\mu_k(\lambda) > 0$ because of the maxmin-characterization of μ_k. Conversely if $\mu_k(\lambda) > 0$, there is a subspace $\hat{V}$ such that $\mu_{\hat{V}}(\lambda) > 0$ and therefore $\lambda > \lambda(\hat{V})$ because of Lemma 4.18. From the minmax-characterization of λ_k it then follows that $\lambda > \lambda_k$.

Since λ_k is the k^{th} eigenvalue in J it follows from the numbering of eigenvalues that $\lambda = \lambda_k$ is equivalent to $\mu_k(\lambda) = 0$, which concludes the proof.

With the last lemma we can proof

LEMMA 5.4
Let $\lambda \in J$ and $x_k(\lambda) \in D$. Then we have

$$\lambda \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} \lambda_k \Leftrightarrow \lambda \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} h_k(\lambda). \tag{5.15}$$

Proof:
We have

$$\begin{aligned} \lambda < \lambda_k &\Leftrightarrow \mu_k(\lambda) < 0 \\ &\Leftrightarrow \langle x_k(\lambda), T(\lambda)x_k(\lambda)\rangle < 0 \\ &\Leftrightarrow \lambda < p(x_k(\lambda)) = h_k(\lambda) \end{aligned}$$

The first equivalence follows from Lemma 5.4. The second equivalence follows from $\mu_k(\lambda) = \langle x_k(\lambda), T(\lambda)x_k(\lambda)\rangle$, and the last equivalence follows from Lemma 4.16. The other cases of the proof are similar.

From Lemma 5.4 it follows directly that λ_k is the only fixpoint of h_k in J and that the iteration (5.13) is aiming in the direction of the fixpoint. But this does not guarantee that one step of (5.13) does not go too far and leads away from the fixpoint.

To establish a statement about the speed of convergence, we need the following lemma.

LEMMA 5.5
Let $T(\lambda)$ be n-times continuous differentiable, $\hat{\lambda}$ a simple eigenvalue of $T(\cdot)$ with corresponding eigenvector $\hat{x}$, $T'(\lambda)$ positive definite in a neighborhood of $\hat{\lambda}$ and $l \in \mathbb{C}^n$ such that $l^H\hat{x} = 1$. Then there exist $\eta > 0$ and n-times continuous differentiable functions

$$\begin{aligned} \mu : (\hat{\lambda} - \eta, \hat{\lambda} + \eta) &\mapsto \mathbb{R} \quad \text{with } \mu(\hat{\lambda}) = 0 \\ x : (\hat{\lambda} - \eta, \hat{\lambda} + \eta) &\mapsto \mathbb{R}^n \quad \text{with } x(\hat{\lambda}) = \hat{x} \end{aligned}$$

such that

$$T(\lambda)x(\lambda) = \mu(\lambda)T'(\lambda)x(\lambda).$$

Proof:

The proof follows from the implicit function theorem. Let

$$\Psi(x, \mu, \lambda) := \begin{bmatrix} (T(\lambda) - \mu T'(\lambda))x \\ l^H x - 1 \end{bmatrix}.$$

Then

$$\Psi(\hat{x}, 0, \hat{\lambda}) = 0.$$

The proof is finished if we can show that $\frac{\partial}{\partial(x,\mu)}\Psi(x, \mu, \lambda)$ is regular in $(\hat{x}, 0, \hat{\lambda})$.

It is

$$\frac{\partial}{\partial(x,\mu)}\Psi(\hat{x},0,\hat{\lambda}) = \begin{pmatrix} T(\hat{\lambda}) & -T'(\hat{\lambda})\hat{x} \\ l^H & 0 \end{pmatrix}.$$

We want to show that from

$$\begin{bmatrix} T(\hat{\lambda}) & -T'(\hat{\lambda})\hat{x} \\ l^H & 0 \end{bmatrix} \begin{pmatrix} t \\ \gamma \end{pmatrix} = 0$$

it follows that $t = 0$ and $\gamma = 0$.

$\gamma = 0$: From the first row we have $T(\hat{\lambda})t = 0$ and therefore $t = \alpha\hat{x}, \alpha \in \mathbb{R}$, since $\hat{\lambda}$ is a simple eigenvalue of $T(\cdot)$. But the second row gives $l^H t = \alpha l^H x = \alpha = 0$ and therefore $\alpha = 0$.

$\gamma \neq 0$: From the first row the following equation is obtained

$$T(\hat{\lambda})t = \gamma T'(\hat{\lambda})\hat{x}.$$

Since $T'(\hat{\lambda})$ is positive definite, we can perform a Cholesky-factorization of $T'(\hat{\lambda})$ and obtain $T'(\hat{\lambda}) = CC^H$. With $\tilde{t} := C^H t$ and $\tilde{x} := C^H\hat{x}$ the above equation can be reduced to

$$C^{-1}T(\hat{\lambda})C^{-H}\tilde{t} = \gamma\tilde{x} \neq 0.$$

Be multiplying the last equation with $C^{-1}T(\hat{\lambda})C^{-H}$ we arrive at

$$(C^{-1}T(\hat{\lambda})C^{-H})^2\tilde{t} = \gamma C^{-1}T(\hat{\lambda})C^{-H}\tilde{x} = 0.$$

Hence, $\tilde{t}$ is a principal eigenvector of order 2, which contradicts the symmetry of $C^{-1}T(\hat{\lambda})C^{-H}$.

If in the previous theorem all occurrences of $T'(\lambda)$ are replaced by the identy matrix I the results stay valid. Hence, we have the same result for the standard eigenvalue problem

$$T(\lambda)x(\lambda) = \mu(\lambda)x(\lambda).$$

A useful result can be obtained for the derivate of

$$T(\lambda)x(\lambda) = \mu(\lambda)T'(\lambda)x(\lambda)$$

at the point $\hat{\lambda}$. The derivate of the previous equation at $\hat{\lambda}$ is

$$T'(\hat{\lambda})x(\hat{\lambda}) + T(\hat{\lambda})x'(\hat{\lambda}) = \mu'(\hat{\lambda})T'(\hat{\lambda})x(\hat{\lambda}). \tag{5.16}$$

By multiplying (5.16) from the left with $\hat{x}^H = x(\hat{\lambda})^H$ we obtain

$$\mu'(\hat{\lambda}) = 1 \tag{5.17}$$

and therefore

$$T(\hat{\lambda})x'(\hat{\lambda}) = 0. \tag{5.18}$$

Since $T(\lambda)$ is continuous differentiable, it follows from Lemma 5.5 that $x_k(\lambda)$ and therefore $p(x_k(\lambda)) = h_k(\lambda)$ is continuous differentiable in a neighborhood of a simple eigenvalue λ_k of $T(\cdot)$. Then we have

$$h_k'(\lambda) = \langle \nabla p(x_k(\lambda)), x_k'(\lambda) \rangle$$

and therefore

$$h_k'(\lambda_k) = 0$$

since $x_k(\lambda_k)$ is a stationary point of $p(x)$. Hence, the fixpoint iteration 5.13 has quadratic convergence near λ_k. If $T'(\lambda)$ is positive definite in J, the linear eigenvalue problem $T(\lambda)x_k(\lambda) = \mu_k(\lambda)x_k(\lambda)$ can be replaced with the generalized eigenvalue problem

$$T(\lambda)x_k(\lambda) = \mu_k(\lambda)T'(\lambda)x_k(\lambda). \tag{5.19}$$

In this case we can even proof cubic convergence if $x_k(\lambda)$ is an eigenvector corresponding to the k^{th} eigenvalue of (5.19). The second derivate of $h_k(\lambda)$ at $\lambda = \lambda_k$ is

$$h_k''(\lambda_k) = -2\frac{< T(p(x_k(\lambda_k)))x_k'(\lambda_k), x_k'(\lambda_k) \rangle}{x_k^H T'(p(x_k(\lambda_k)))x_k}.$$

Because of equation 5.18

$$T(p(x_k(\lambda_k)))x_k'(\lambda_k) = 0$$

and thus $h_k''(\lambda_k) = 0$. This means that we have even cubic convergence near a simple eigenvalue λ_k.

We will later use the guarded iteration as inner iteration in the Jacobi-Davidson method for nonlinear eigenvalue problems. The fast convergence of this method and the ability to directly iterate for an eigenvalue with a given number let this method be a very suitable inner iteration for projection methods.

5.2 Nonlinear Rational Krylov

The Nonlinear Rational Krylov method is a projection method for nonlinear eigenvalue problems where similar to Arnoldi a subspace is built up from which approximations for the eigenvalues are extracted. Since the Nonlinear Rational Krylov method is based on a number of linearizations of the nonlinear eigenvalue problem it is most suitable for weakly nonlinear problems in which the nonlinearity has no strong influence - for example in vibration problems with weak damping.

The basic idea is to interpolate the nonlinear operator $T(\lambda)$ between two points σ and μ. Then we have

$$T(\lambda) \approx \tilde{T}(\lambda) := \frac{\lambda - \sigma}{\mu - \sigma} T(\mu) + \frac{\mu - \lambda}{\mu - \sigma} T(\sigma) \tag{5.20}$$

We iterate for an eigenvalue of $T(\cdot)$ by letting σ be fixed and updating μ in every step with the solution of the linear problem $\tilde{T}(\lambda)x = 0$. This can be seen as a secant iteration for $T(\lambda)x = 0$. The linear eigenvalue problem $\tilde{T}(\lambda)x = 0$ can be rewritten as

$$\left[\frac{\mu_{k+1} - \sigma}{\mu_k - \sigma} T(\mu) + \frac{\mu_k - \mu_{k+1}}{\mu_k - \sigma} T(\sigma) \right] x_k = 0 \tag{5.21}$$

with $\lambda = \mu_{k+1}$. Let

$$\theta = \frac{\mu_{k+1} - \mu_k}{\mu_{k+1} - \sigma}.$$

Then equation (5.21) is equivalent to

$$\left[T(\sigma)^{-1} T(\mu_k) - \theta I \right] x_k = 0 \tag{5.22}$$

We have the following first algorithm

Algorithm 7 Regula Falsi for Nonlinear Eigenvalue Problems

1: Start with a pole σ and an approximation μ_1 for an eigenvalue of (4.1).
2: **for** $k = 1, \ldots$ until convergence **do**
3: Solve

$$T(\sigma)^{-1} T(\mu) x = \theta x$$

4:

$$\mu_{k+1} = \mu_k + \frac{\theta}{1 - \theta} (\mu_k - \sigma) \tag{5.23}$$

5: **end for**

Let us review the update (5.23). μ_{k+1} is an approximation for an eigenvalue of $T(\cdot)$. Hence, eigenvalues of $T(\cdot)$ far away from the pole and the current approximation μ_k correspond to values of θ near 1. Values of θ near 0 lead to a small update of μ_k and we thus arrive at eigenvalues of $T(\cdot)$ near μ_k.

We now want to combine Algorithm 7 with the Arnoldi method to solve the linear eigenvalue problem in every step. As it was explained in chapter 3.2 the basic Arnoldi recursion (see equation (3.5)) for the eigenvalue problem (5.22) is

$$T(\sigma)^{-1} T(\mu) V_j = V_j H_j + h_{j+1,j} v_{j+1} e_j^H.$$

or with $H_{j+1,j}$ denoting the $j+1 \times j$ Hessenberg Matrix consisting of H_j and the element $h_{j+1,j}$.

$$T(\sigma)^{-1} T(\mu) V_j = V_{j+1} H_{j+1,j}. \tag{5.24}$$

Instead of the solution of the eigenvalue problem (5.22) we use the Ritz-Values obtained from the solution of $H_m s = \theta s$ in every step of Algorithm 7. The problem which still has to be solved is how to describe an Arnoldi recursion for the nonlinear operator $T(\sigma)^{-1}T(\mu)$ in which μ changes in every step. In the linear case described in chapter 3.2 the current subspace is extended with the residual $r = (A - \theta)V_m s$ where (θ, s) is an eigenpair of $H_{m,m}$. This residual is orthogonal to the current subspace V_m. After extending V_m with r to V_{m+1} the matrix A is projected on this new subspace. We now want to achieve something similar for the nonlinear case.

If Arnoldi and Regula Falsi are strictly combined we obtain the following update formula

$$\begin{aligned} T(\sigma)^{-1}T(\mu_j)V_j &= V_{j+1}H^j_{j+1,j} \\ H^j_{j,j}s &= \theta_j s \\ r &= T(\sigma)^{-1}T(\mu_j)V_j s \\ v_{j+1} &= r - V_j V_j^H r \\ \mu_{j+1} &= \mu_j + \frac{\theta}{1-\theta}(\mu_j - \sigma). \end{aligned}$$

To let this update formula be more Arnoldi like, Hager and Ruhe and Wiberg introduced an update for $H^j_{j+1,j}$ which does not depend on the explicit projection

$$H^j_{j+1,j} = V_{j+1}T(\sigma)^{-1}T(\mu_j)V_j, \tag{5.25}$$

since it is in general not a Hessenberg matrix and may be in some cases expensive to update.

The idea of the simplified update formula is to let only the last column of the new Hessenberg matrix be true orthogonalization coefficients and to use an update formula of the type $\tilde{H} = \alpha H - \beta I$ for the rest of the Hessenberg matrix. Hence, we use $\tilde{H}^j_{j+1,j} = \begin{bmatrix} \tilde{H}^j_{j,j-1} & h_j \\ 0 & \end{bmatrix} = \begin{bmatrix} \tilde{H}^j_{j+1,j-1} & h_j \end{bmatrix}$, where h_j are the true orthogonalization coefficients $h_j = V^H_{j+1}T(\sigma)^{-1}T(\mu_j)v_j$ and $\tilde{H}^j_{j,j-1}$ is the updated $\tilde{H}^{j-1}_{j,j-1}$.

Assume that we already have

$$T(\sigma)^{-1}T(\mu_{j-1})V_{j-1} = V_j H^{j-1}_{j,j-1}$$

in step $j-1$ with the true projected matrix $H^{j-1}_{j,j-1}$. In step j the matrix $\tilde{H}^j_{j+1,j}$ shall be used. We obtain

$$T(\sigma)^{-1}T(\mu_j)V_j = V_{j+1}\tilde{H}^j_{j+1,j} + R_j$$

where R_j is the error matrix due to the usage of the simplified update. Then

$$\begin{aligned} R_j &= T(\sigma)^{-1}T(\mu_j)V_j - V_{j+1}\left[\alpha_j H_{j+1,j-1}^{j-1} - \beta_j I_{j+1,j-1} \quad h_j\right] \\ &= T(\sigma)^{-1}T(\mu_j)V_j - \left[\alpha_j V_j H_{j,j-1}^{j-1} - \beta_j V_{j-1} \quad V_{j+1}h_j\right] \\ &= \big[T(\sigma)^{-1}T(\mu_j)V_{j-1} - \alpha_j T(\sigma)^{-1}T(\mu_{j-1})V_{j-1} + \beta_j V_{j-1} \ldots \\ &\quad T(\sigma)^{-1}T(\mu_j)v_j - V_{j+1}h_j\big] \\ &= \left[(T(\sigma)^{-1}T(\mu_j) - \alpha_j T(\sigma)^{-1}T(\mu_{j-1}) + \beta_j I)V_{j-1} \quad 0\right]. \end{aligned}$$

If $T(\sigma)^{-1}T(\mu_j) \approx \alpha_j T(\sigma)^{-1}T(\mu_{j-1}) - \beta_j I$ we have $R_j \approx 0$. A Lagrange-Interpolation of $T(\sigma)^{-1}T(\mu_j)$ between the identity matrix I and the matrix $T(\sigma)^{-1}T(\mu_{j-1})$ yields

$$T(\sigma)^{-1}T(\mu_j) \approx \frac{\mu_j - \sigma}{\mu_{j-1} - \sigma} T(\sigma)^{-1}T(\mu_{j-1}) - \frac{\mu_j - \mu_{j-1}}{\mu_{j-1} - \sigma} I,$$

from which the coefficients α_j and β_j can be identified. We finally arrive at

$$\tilde{H}_{j,j-1}^{j} = \frac{\mu_j - \sigma}{\mu_{j-1} - \sigma} \tilde{H}_{j,j-1}^{j-1} - \frac{\mu_j - \mu_{j-1}}{\mu_{j-1} - \sigma} I_{j,j-1} \tag{5.26}$$

Now we have an easy to compute approximative update formula. Let us formulate a first version of the Nonlinear Rational Krylov Algorithm

Algorithm 8 Nonlinear Rational Krylov

1: Start with v_1, $\|v_1\| = 1$, σ, μ_1
2: **for** $j = 2, \ldots$ until convergence **do**
3: $r = T(\sigma)^{-1}T(\mu_j)v_j$
4: $h_j = V_j^H r$, $r = r - V_j h_j$
5: $h_{j+1,j} = \|r\|$, $v_{j+1} = r/\|r\|$
6: Solve the Eigenvalue Problem $H_{j,j}^{j} S = S\Theta$
7: Select the $\tilde{\theta}$ having the smallest modulus and the corresponding eigenvector $\tilde{s}$
8: $x = V_j \tilde{s}$.
9:

$$\mu_{j+1} = \mu_j + \frac{\theta}{1-\theta}(\mu_j - \sigma)$$

10: Test for convergence $\|T(\mu_{j+1})x\|/\|x\| \leq \text{tol}$
11: Update σ, perform a restart if necessary
12: Use the update-formula

$$H_{j,j-1}^{j} = \frac{\mu_j - \sigma}{\mu_{j-1} - \sigma} H_{j,j-1}^{j-1} - \frac{\mu_j - \mu_{j-1}}{\mu_{j-1} - \sigma} I_{j,j-1}$$

or compute

$$H_{j+1,j}^{j+1} = V_{j+1,j}^H T(\sigma)^{-1}T(\mu_{j+1})V_j$$

13: **end for**

Practical implementation details for this algorithm can be found in [10]. This iteration is already very similar to the Rational Krylov Algorithm for linear eigenvalue problems. If the linear eigenvalue problem $T(\lambda) = \lambda B - A$ is used, we have

$$T(\sigma)^{-1}T(\mu) = (\sigma B - A)^{-1}(\mu B - A).$$

An exact instance of Rational Krylov is obtained if $\mu \to \infty$ and H is exactly computed in every step. With the choice $\mu \to \infty$ we obtain the operator

$$T(\sigma)^{-1}T(\mu) = (\sigma B - A)^{-1}B.$$

In this case we just have an instance of Algorithm 4 if updates of σ are incorporated. There is still one obstacle. For linear eigenvalue problems the residual $r = T(\mu)x$ is always orthogonal to the current subspace V_j. This is in general not true for nonlinear eigenvalue problems. Hager, Ruhe and Wiberg proposed a heuristical approach to obtain the orthogonality of the residual. It is based on the idea to operate on the last column of $H^j_{j+1,j}$ until the residual is orthogonal to the current subspace. Instead of the orthogonalization in line 4 of Algorithm 8 and the update of H in line 11 an inner iteration is used.

Algorithm 9 NLRKS with residual iteration

1: Start with v_1, $\|v_1\| = 1$, σ, μ_1, $j = 1$
2: $h_j = 0$, $s = e_j$; $x = v_j$
3: $r = T(\sigma)^{-1}T(\mu_j)v_j$
4: $k = V_j^H r$
5: **if** $\|k_j\| \leq toln$ **then**
6: goto 15
7: **end if**
8: $r = r - V_j k_j$
9: $h_j = h_j + k_j s(j)^{-1}$
10: Solve the Eigenvalue Problem $H_{j,j}S = S\Theta$
11: Select the $\tilde{\theta}$ having the smallest modulus and the corresponding eigenvector $\tilde{s}$; $s = \tilde{s}$.
12: $x = V_j\tilde{s}$.
13:
$$\mu_{j+1} = \mu_j + \frac{\theta}{1-\theta}(\mu_j - \sigma)$$
14:
$$H^j_{j,j-1} = \frac{\mu_j - \sigma}{\mu_{j-1} - \sigma}H^{j-1}_{j,j-1} - \frac{\mu_j - \mu_{j-1}}{\mu_{j-1} - \sigma}I_{j,j-1}$$
15: $h_{j+1,j} = \|r\|/s_j$
16: $v_{j+1} = r/\|r\|$, $j = j + 1$
17: Test for convergence $\|T(\mu_{j+1}x\|/\|x\| \leq$ tol
18: Update σ, perform a restart if necessary
19: If more eigenvalues are wanted, choose next $\tilde{\theta}$ and corresponding s, goto 12

The tolerance *toln* denotes the threshold for the orthogonality of r towards V_j. Hints for practical implementations and examples can be found in ([11]).

The update of the pole σ is performed similar to the linear case which was described in 3.3. Let σ, μ be the current pole respectively shift and $\tilde{\sigma}, \tilde{\mu}$ the new pole and shift. Then a Lagrange-Interpolation yields

$$\begin{aligned} T(\mu) &= \frac{\mu-\tilde{\sigma}}{\tilde{\mu}-\tilde{\sigma}}T(\tilde{\mu}) + \frac{\tilde{\mu}-\mu}{\tilde{\mu}-\tilde{\sigma}}T(\tilde{\sigma}) \\ T(\sigma) &= \frac{\sigma-\tilde{\sigma}}{\tilde{\mu}-\tilde{\sigma}}T(\tilde{\mu}) + \frac{\tilde{\mu}-\sigma}{\tilde{\mu}-\tilde{\sigma}}T(\tilde{\sigma}) \end{aligned}$$

Substituting this into 5.24 and rearranging the terms yields

$$T(\tilde{\sigma})^{-1}T(\tilde{\mu})V_{j+1}K_{j+1,j} = V_{j+1}L_{j+1,j} \tag{5.27}$$

with

$$\begin{aligned} K_{j+1,j} &= \frac{\mu-\tilde{\sigma}}{\tilde{\mu}-\tilde{\sigma}}I_{j+1,j} - \frac{\sigma-\tilde{\sigma}}{\tilde{\mu}-\tilde{\sigma}}H_{j+1,j} \\ L_{j+1,j} &= \frac{\tilde{\mu}-\tilde{\sigma}}{\tilde{\mu}-\tilde{\sigma}}H_{j+1,j} - \frac{\tilde{\mu}-\mu}{\tilde{\mu}-\tilde{\sigma}}I_{j+1,j} \end{aligned} \tag{5.28}$$

Now we can proceed exactly as described in chapter 3.3 to transform (5.27) into an Arnoldi recursion of the form (5.24).

A Lock-And-Purge operation to reduce the subspace can also be performed similar to the linear case. Implementation details can be found in ([11]).

The NLRKS method has several drawbacks which need further research. The main problem is that the structure of the nonlinear eigenvalue problem is completely destroyed by operating on $T(\sigma)^{-1}T(\mu)$ instead of $T(\mu)$. For example symmetry properties of $T(\cdot)$ are generally lost and if one is searching only for real eigenvalues it can not be guaranteed that the updates for μ obtained from H_j are real. The next problem is that the motivation for the residual iteration r is still heuristically and it is not yet clear under which circumstances it converges. The extension of Rational Krylov for nonlinear eigenvalue problems is certainly an interesting approach as it is an idea to extend the approximation of the spectrum with rational functions in the linear case to arbitrary nonlinear problems. It was also the first approach from which we know that uses projection methods to solve large sparse nonlinear eigenvalue problems without several LU factorizations in every step. But there is still some research necessary to fully understand the convergence behaviour of this algorithm.

5.3 A Jacobi-Davidson type method

Now we want to introduce a new approach for sparse symmetric nonlinear eigenvalue problems which is based on the Jacobi-Davidson method (cf. [3]). Present solution methods have the problem that they need several decompositions of the operator $T(\cdot)$ in every step to solve the nonlinear eigenvalue problem (4.1). Hence, they are not suitable for the extraction of several eigenvalues from large sparse eigenvalue problems which appear in many applications. A first idea to overcome this problem was the NLRKS method discussed in the previous section. However, this method seems to be only suitable for weakly nonlinear problems, e.g. vibration problems with small damping, because it heavily depends on linearizations to apply Rational Krylov to the nonlinear problem. Our approach is different. We try to retain the structure of the nonlinear eigenvalue problem by constructing a subspace with the Jacobi-Davidson approach and projecting the nonlinear eigenvalue problem onto this subspace. Then the projected problem is solved with methods for dense nonlinear eigenvalue problems. A similar Jacobi-Davidson type approach was already introduced by Sleijpen and van der Vorst for quadratic eigenvalue problems (cf. [30]). But they did not work out strategies to find several eigenvalues with the Jacobi-Davidson approach without the problem of convergence against old eigenvalues that were already extracted in previous steps of the algorithm. This shall be done here.

The idea of Jacobi-Davidson for the nonlinear case is very similar to the linear case. But instead of the correction equation 3.18 we use the following equation to extend the actual search subspace spanned by the columns of v.

$$(I - \frac{pu^H}{u^H p})T(\sigma)(I - \frac{uu^H}{u^H u})t = -r, \quad t \perp u, \tag{5.29}$$

Here, (σ, u) is the current approximation for an eigenpair of $T(\cdot)$, p denotes the vector $p = T'(\sigma)u$ and r denotes the residual $r = T(\sigma)u$. If $T(\lambda) = \lambda I - A$ we arrive at the correction equation 3.18 for the linear case. This extension was already proposed by Sleijpen and van der Vorst in [30].

We want to perform a similar analysis for equation (5.29) as in chapter 3.4. The correction equation (5.29) can be written as

$$T(\sigma)t - \alpha p = -r$$

where α must be chosen such that $t \perp u$. The last equation leads to

$$t = -u + \alpha T(\sigma)^{-1}p$$

since $r = T(\sigma)u$. By substituting p with $T'(\sigma)^{-1}u$ we arrive at

$$t = -u + \alpha T(\sigma)^{-1}T'(\sigma)u.$$

Like in the linear case the vector t is orthogonalized to the current subspace V which contains u. Hence, we can write

$$t = \alpha T(\sigma)^{-1}T'(\sigma)u,$$

which is exactly one step of inverse iteration for nonlinear eigenvalue problems. If $\tilde{t}$ denotes the direction of t orthogonal to the current subspace spanned by V_k, hence $\tilde{t} = t - VV^H t$, we have the following statement

$$T(\sigma)^{-1}T'(\sigma)u \in span(\{V_{k+1}\}) = span(\{[V_k, \tilde{t}]\}),$$

because the basis V_k of the current search subspace is extended with $\tilde{t}$ in the Jacobi-Davidson iteration. Therefore like in the linear case the new search subspace V_{k+1} contains the approximation for an eigenvector which is obtained from inverse iteration. Inverse iteration for nonlinear eigenvalue problems converges cubically in the symmetric case if the Rayleigh functional is used to update the eigenvalue approximation. Hence, we can expect asymptotically cubic convergence also in the nonlinear version of Jacobi-Davidson if the correction equation is solved exactly and the projected nonlinear eigenvalue problem is solved in a suitable way.

For the further analysis we want to write down a first version of the nonlinear Jacobi-Davidson method.

Algorithm 10 Nonlinear Jacobi-Davidson

1: Start with $V = v_1/\|v_1\|$
2: $n = 1$, $k = 1$
3: **while** $n \leq$ Number of wanted Eigenvalues **do**
4: Compute the wanted eigenvalue θ and the corresponding eigenvector y of $V^H T(\lambda)Vy = 0$.
5: $\sigma_k = \theta$, $u_k = Vy$, $r_k = T(\sigma_k)u_k/\|u_k\|$
6: **if** $\|r_k\| < \epsilon$ **then**
7: PRINT σ_k,u_k
8: $n = n + 1$
9: GOTO 3
10: **end if**
11: $p_k = T'(\sigma_k)u_k$
12: Find an approximate solution for the correction equation

$$(I - \frac{p_k u_k^H}{u_k^H p_k})T(\sigma_k)(I - \frac{u_k u_k^H}{u_k^H u_k})t = -r_k, \quad t \perp u_k,$$

13: $t = t - VV^H t$,$\tilde{v} = t/\|t\|$, $V = [V, \tilde{v}]$
14: If necessary perform a restart to reduce the size of the subspace V.
15: $k = k + 1$
16: **end while**

There are no great differences to the linear version of Jacobi-Davidson described in Algorithm 5. The main difference is that deflation strategies are not incorporated in this algorithm. This is due to the fact that there do not yet exist any deflation strategies for nonlinear eigenvalue problems. In the linear Hermitian case it is easy. There the eigenvectors are orthogonal to each other and therefore eigenvectors to already found eigenvalues can easily be extracted from the current search subspace without loosing information for new eigenvectors. In the nonlinear case we only have orthogonality according to a form which is in general not even bilinear (cf. [8]).

So we need other methods than deflation to ensure that Algorithm 10 does not converge to already found eigenvalues. In the symmetric case we can find a solution by exploiting the minmax properties of symmetric eigenvalue problems. The minmax theorems described in chapter 4.5 give a numbering of the eigenvalues which is used by the guarded iteration to directly iterate for the k^{th}-eigenvalue of $T(\cdot)$. The idea is now to use the guarded iteration to find the k^{th}-eigenvalue of the projected problem $V^H T(\lambda) V y = 0$ in the course of Algorithm 10 and to use this approximation to find a suitable update with the correction equation. Thus, we combine the Jacobi-Davidson type outer iteration with an inner iteration for the k^{th} eigenvalue of the projected problem. Since the projected eigenvalue problem is usually of small dimension, there is no great effort in finding the k^{th} eigenvalue of the projected problem with the guarded iteration. So the Jacobi-Davidson iteration aims directly towards the k^{th} eigenvalue and the corresponding eigenvector of the unprojected problem $T(\lambda)x = 0$. This solves the problem of deflation, because if the k^{th} eigenvalue of $T(\cdot)$ is found, we start iterating for the eigenvalue with the number $k + 1$.

Now we want to discuss how to efficiently apply an iterative Krylov subspace solver to find approximate solutions of the correction equation (5.29). The method proposed here is almost identical to the linear case which is already described in [29].

We want to apply a Krylov subspace solver for the correction equation (5.29). For the linear system $Ax = b$ a Krylov subspace solver takes approximations for the solution x from the Krylov subspace

$$K_m(A, v) \equiv \text{span}\{v, Av, A^2v, \ldots, A^{m-1}v\}$$

Simply building up a Krylov subspace for the operator

$$\tilde{A} = (I - \frac{p_k u_k^H}{u_k^H p_k}) T(\sigma_k) (I - \frac{u_k u_k^H}{u_k^H u_k})$$

must fail, since $\tilde{A}$ maps onto the subspace $\{T'(\sigma_k)u_k\}^\perp$. But we are looking for solutions of the correction equation in the subspace $\{u_k\}^\perp$. Therefore some sort of preconditioner $\tilde{K}$ is required such that the operator $\tilde{K}^{-1}\tilde{A}$ from the preconditioned correction equation

$$\tilde{K}^{-1}\tilde{A}t = -\tilde{K}r \tag{5.30}$$

maps from $\{u_k\}^\perp$ to $\{u_k\}^\perp$. Then for the Krylov subspace $K_m(\tilde{K}^{-1}\tilde{A}, v)$ we have

$$K_m(\tilde{K}^{-1}\tilde{A}, v) \subset \{u_k\}^\perp,$$

if $v \in \{u_k\}^\perp$. Hence, from $K_m(\tilde{K}^{-1}\tilde{A}, v)$ an approximation $t \in \{u_k\}^\perp$ for the solution of (5.29) can be extracted.

Let K be a an approximation for the matrix $T(\sigma_k)$ for which an LU factorization can easily be obtained. Then we define $\tilde{K}$ as

$$\tilde{K} := (I - \frac{p_k u_k^H}{u_k^H p_k})K(I - \frac{u_k u_k^H}{u_k^H u_k}).$$

Hence, $\tilde{K}$ maps from $\{u_k\}^\perp$ to $\{T'(\sigma_k)u_k\}^\perp$. Therefore the operator $\tilde{K}^{-1}\tilde{A}$ maps from $\{u_k\}^\perp$ to $\{u_k\}^\perp$ and we can apply a Krylov subspace solver to solve the preconditioned system (5.30).

The question remains how to apply the operator $\tilde{K}^{-1}\tilde{A}$ to a vector v in the Krylov subspace method. Sleijpen and van der Vorst demonstrated an easy way to perform this operation which does not have significantly more cost than solving a preconditioned system which does not involve projections. Consider that $v \in \{u_k\}^\perp$, then the multiplication

$$y = \tilde{K}^{-1}\tilde{A}v$$

can be performed in two steps. First multiply v by $\tilde{A}$ which yields

$$\tilde{y} = (I - \frac{p_k u_k^H}{u_k^H p_k})T(\sigma_k)v,$$

and then solve

$$\tilde{K}y = \tilde{y}, \quad y \perp u_k.$$

This equation can be rewritten as

$$Ky - \alpha p_k = \tilde{y},$$

where α is determined from the condition that $y \perp u_k$. Thus, we finally obtain

$$y = K^{-1}\tilde{y} - \frac{u_k^H K^{-1}\tilde{y}}{u_k^H K^{-1} p_k} K^{-1} p_k$$

Therefore the approximative solution of (5.29) with a preconditioned Krylov-solver requires one application of the preconditioner K^{-1} to get $K^{-1}p_k$, and one more application of K^{-1} in each step to obtain $K^{-1}\tilde{y}$. Hence, for the correction equation which involves projection only one more application of K^{-1} is needed compared to a linear system which does not involve projections.

Since the linear equation (5.30) is in general not symmetric, we use GMRES as Krylov subspace solver. As start vector for GMRES, the vector $t = 0$ is used. The great advantage of Jacobi-Davidson is that only some steps of GMRES are necessary to obtain

a suitable correction. We could observe this behavior also in the case of a nonlinear eigenvalue problem. Hence, this method allows us to extract several eigenvalues of $T(\cdot)$ without the need to perform expensive LU decompositions of $T(\cdot)$.

At last we want to give a possibility how restarts can be performed in the course of Algorithm 10.

From the proof of Theorem 4.15 we know that the minimum in the characterization

$$\lambda_\ell = \min_{V \in H_{\ell,D}} \sup_{x \in V \cap D} p(x)$$

is reached for the invariant subspace $\bar{V}$ of eigenvectors to the ℓ largest eigenvalues of $T(\lambda_\ell)$. Let $\bar{V}$ also denote the matrix of the eigenvectors to the ℓ largest eigenvalues of $T(\lambda_\ell)$. Then the ℓ-th eigenvalue of

$$\bar{V}^H T(\lambda) \bar{V} = 0$$

is λ_ℓ and any subspace containing $\bar{V}$ yields a perfect restart. Let V be our current basis for the search subspace. Then we use as approximation for $\bar{V}$ the subspace $\tilde{V} = VS$, where S is the matrix of eigenvectors to the ℓ largest eigenvalues of

$$V^H T(\sigma_k) V x = \mu_x$$

and continue the Jacobi-Davidson iteration with the reduced subspace $\tilde{V}$. In one example it occured that the iteration needed unusually many steps after a restart to find the next eigenvalue. Therefore a relaxed strategy may be sometimes more suitable in which not only the eigenvectors to the ℓ largest eigenvalues but to the $\tilde{\ell} = \ell + l$ largest eigenvalues are taken. In our tests $l = 3$ was a reasonable choice for all examples.

If $T'(\lambda)$ is positive definite then we replace S in the last paragraph by the matrix containing an orthogonal basis for ℓ or $\tilde{\ell}$ eigenvectors corresponding to the largest ℓ or $\tilde{\ell}$ eigenvalues of the generalized matrix eigenvalue problem

$$V^H T(\sigma_k) V x = \mu V^H T'(\sigma_k) V x.$$

This modification of the restart procedure can be motivated by the linear maxmin theory. λ_ℓ is an ℓ^{th} eigenvalue and x_ℓ a corresponding eigenvector of $T(\cdot)$ if and only if $\mu_\ell = 0$ is the ℓ-largest eigenvalue of the linear problem

$$T(\lambda_\ell) x = \mu x$$

and x_ℓ is a corresponding eigenvector.

If σ_k is a good approximation to λ_ℓ, then

$$T(\lambda_\ell) \approx T(\sigma_k) - \eta T'(\sigma_k)$$

is a first order approximation, and we can approximate (4.1) by the generalized eigenproblem

$$(T(\sigma_k) - \eta T'(\sigma_k))x_\ell = 0. \tag{5.31}$$

The ℓ largest eigenvalue η_ℓ is near 0, if σ_k is a good approximation to λ_ℓ, and it can be characterized by a maxmin principle

$$0 \approx \eta_\ell = \max_{V \in S_\ell} \min_{v \in V \setminus \{0\}} \frac{v^H T(\sigma_k) v}{v^H T'(\sigma_k) v}.$$

The maximum is attained by the subspace spanned by the eigenvectors corresponding to the ℓ largest eigenvalues of (5.31) which motivates the choice of S.

In the next chapter we will demonstrate the capability of this new approach for symmetric eigenvalue problems. For nonsymmetric problems there occurs the problem that the eigenvalues are not ordered but lie somewhere in the complex plane. However, for the special case of vibration problems involving small damping we could successfully extend our approach by exploiting the fact that eigenvectors of the damped and undamped problem are often very near. For arbitrary nonlinear eigenvalue problems we are currently working on an approach which includes guarding the method of successive linear approximations. In chapter 7 we will present these ideas together with numerical examples for a nonsymmetric problem.

Chapter 6

Numerical tests

In this chapter we want to present an application of the new method introduced in chapter 5.3 to a symmetric rational eigenvalue problem from fluid-solid vibration.

6.1 The definition of the problem

We want to describe the eigenvalues of a rational eigenvalue problem governing free vibrations of a tube bundle immersed in a slightly compressible fluid. We make following simplifying assumptions. The tubes are assumed to be rigid and assembled parallel inside the fluid. They are mounted in a way such that they can vibrate transversally, but not perpendicular to their sections. The fluid is assumed to be contained in an infinite long cavity. Each tube is supported by an independent system of springs, which simulates the specific elasticity of each tube. Hence, three-dimensional effects can be neglected and we can study the problem in any transversal direction of the cavity. We consider only small vibrations of the fluid around the state of rest and assume that the fluid is irrotational.

The mathematical description is the following (cf. [5],[17]) Let $\Omega \in \mathbb{R}^2$ (the section of the cavity) be an open bounded set with locally Lipschitz continuous boundary Γ. We assume that there exists a family $\Omega_j \neq \emptyset, j = 1, \ldots, K$, (the sections of the tubes) of simply connected open sets such that $\bar{\Omega}_j \subset \Omega$ for every j, $\bar{\Omega}_j \cap \bar{\Omega}_i = \emptyset$ for $j \neq i$ and each Ω_j has locally Lipschitz continuous boundary Γ_j. With these notations we set $\Omega_0 := \Omega \backslash \bigcup_{j=1}^{K} \Omega_j$. Then the boundary of Ω_0 consists of $K+1$ connected components which are Γ and $\Gamma_j, j = 1, \ldots, K$.

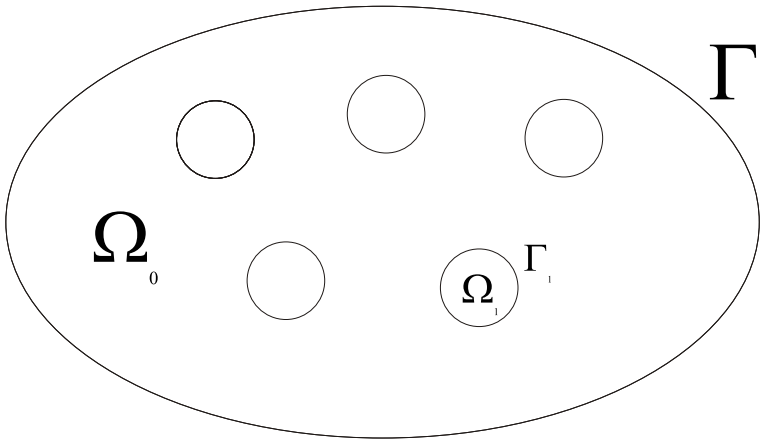

Figure 6.1: The problem of tubes immersed in a confined fluid.

The equations governing the problem are (cf. [17]):

$$\begin{aligned} \Delta\Phi &= -\frac{\omega^2}{c^2}\Phi \quad \text{in } \Omega_0 \\ \frac{\partial\Phi}{\partial n}(x) &= \rho_0 \frac{\omega^2}{k_i - m_i\omega^2}\vec{n}_x \cdot \int_{\Gamma_i} \Phi(y)\vec{n}_y \mathrm{d}\Gamma_y \quad \text{on } \Gamma_i, i = 1, \ldots, K \\ \frac{\partial\Phi}{\partial n}(x) &= 0 \quad \text{on } \Gamma \end{aligned} \tag{6.1}$$

The variables in these equations are:

ρ_0	The specific density of the fluid
m	The tube mass
k	The stiffness of the spring system
Φ	The potential of the velocity
ω	The angular frequency of the structures
$\vec{n}$	The outward unit normal vector

We want to find a weak formulation of the problem. Let $H^1(\Omega_0) := \{u \in L^2(\Omega_0) : \nabla u \in L^2(\Omega_0)^2\}$ be the standard Sobolev space with the usual scalar product. By multiplying 6.1 with a function $v \in H^1(\Omega_0)$, integrating by parts and considering the boundary conditions we obtain the weak formulation as

Find $\lambda \in \mathbb{R}$ *and* $u \in H^1(\Omega_0)$ *such that for every* $v \in H^1(\Omega_0)$

$$c^2 \int_{\Omega_0} \nabla u \cdot \nabla v dx = \lambda \int_{\Omega_0} uv \; dx + \sum_{j=1}^{K} \frac{\lambda\rho_0}{k_j - \lambda m_j} \int_{\Gamma_j} u\vec{n} \; ds \cdot \int_{\Gamma_j} v\vec{n} \; ds. \tag{6.2}$$

In [36] Voss applied the minmax theory for nonoverdamped eigenvalue problems to characterize the eigenvalues of this problem.

For our numerical tests we used the following data: Ω is the ellipse with center $(0,0)$ and length of semiaxes 8 and 4. $\Omega_j, j = 1,\ldots,9$ are circles with radius 0.3 and centers $(-4,-2)$, $(0,-2)$, $(4,-2)$, $(-5,0)$, $(0,0)$, $(5,0)$, $(-4,2)$, $(0,2)$ and $(4,2)$. All constants in the formulation of problem (6.2) are set to 1.

By a finite-element discretization of problem (6.2) one obtains the rational matrix eigenvalue problem

$$T(\lambda)x := -Kx + \lambda Mx + \frac{\lambda}{1-\lambda}Cx = 0. \tag{6.3}$$

The matrix C collects the contributions of all tubes. K is the stiffness matrix and M the mass matrix of the system. All matrices are symmetric. K and C are positive semidefinite and M is positive definite. The dimension of the matrices in our numerical tests is 36040.

The problem has 28 eigenvalues in $J_1 := (0,1)$ and 20 eigenvalues in $J_2 := (1,3)$. All tests were performed on a Pentium 4 with 2 GHz and 512 MByte Ram. The test environment was Matlab 6.1 running on Windows 2000.

6.2 Preconditioning

As preconditioner for the correction equation (5.29) we used a few numbers of LU decompositions of $T(\sigma_k)$ for different values of σ_k. To guard the quality of the LU decomposition corresponding to a given σ_k for the values $\sigma_j, j = k+1,\ldots$ we observed the number of iterations needed to obtain a given accuracy with GMRES for the correction equation. If GMRES took too many steps, it was time to update the preconditioner. This strategy has shown to be very effective for this problem, since performing some few LU decompositions is not too expensive for the example problem and the convergence behaviour is drastically improved.

To be precise, for the first test we used an accuracy of 1E-2 for the residual in the GMRES iteration. If this accuracy was not obtained within 5 steps, a new LU decomposition was performed. The update of the LU decompositions first started after the first two eigenvalues converged to give the algorithm time to build up a sufficient subspace. We allowed a maximum number of 10 GMRES steps. With this preconditioner we obtained the convergence history for the 28 eigenvalues in $(0,1)$ shown in figure 6.2. An approximation (λ, x) for an eigenpair was accepted as eigenpair when its residual $\|T(\lambda)x\|/\|x\|$ was below 1E-13. In this case Algorithm 10 needed 14 LU-decompositions and 363 GMRES steps for all 28 eigenvalues. The computation was finished after 15 minutes. This also shows that the correction equation needs only be solved with a small accuracy to obtain a good search direction. Solving the correction equation with higher accuracy does not automatically mean an advantage. In one test

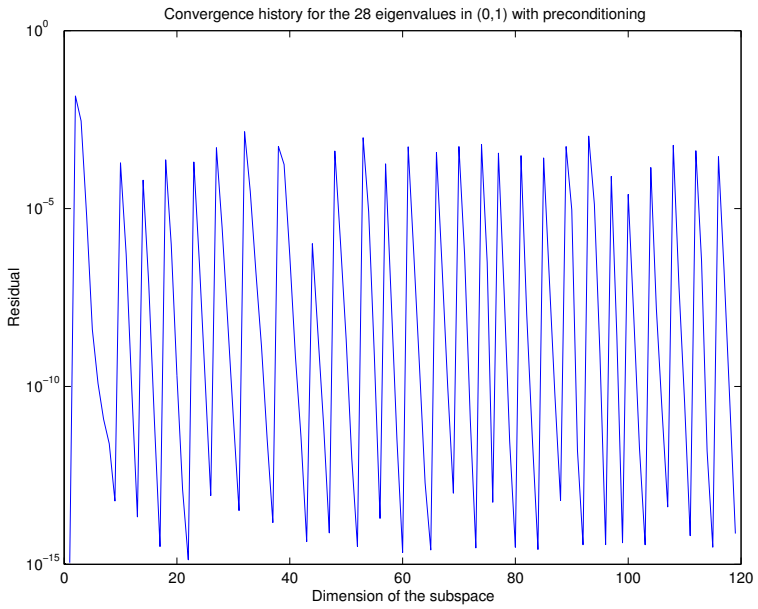

Figure 6.2: Tolerance for GMRES 1E-2. Update of the LU-Decomposition after 5 GMRES steps

we set the tolerance for GMRES to 1E-5. The dimension of the subspace to find all eigenvalues decreased to 98. But the number of GMRES steps increased to 428 and now 24 LU-decompositions were needed with our update strategy.

We also used ILU with a drop tolerance of 1E-3 as preconditioner. Since in this case the residual of the GMRES iteration never reached 1E-2 within 10 steps we used the strategy that for every eigenvalue one new ILU-decomposition was performed. This strategy brought no advantage in computation time. But now a subspace dimension of 209 was reached until all eigenvalues were found and the algorithm needed 2080 GMRES steps since in every step of the algorithm 10 steps of GMRES were performed. But this result shows that LU-decompositions are not necessary as preconditioner. This is a great advantage compared to methods for nonlinear eigenvalue problems which depend on LU decompositions.

Without any preconditioning our test problem did not converge. After 100 steps of Algorithm 10 we obtained the convergence history shown in figure 6.3. The first eigenvalue is found almost immediately. This is due to the fact that we start near the zero eigenvalue and use the eigenvector to the zero eigenvalue of the linear problem $Kx = \lambda Mx$ as start vector for the Jacobi-Davidson iteration. This is also the eigenvector of the nonlinear problem $Kx = \lambda Mx + \lambda/(1-\lambda)Cx$ corresponding to the zero eigenvalue of the nonlinear problem. But the convergence for the second eigenvalue is extremely slow. Hence, some sort of preconditioner is always advisable for the correction equation if the problem is not well conditioned.

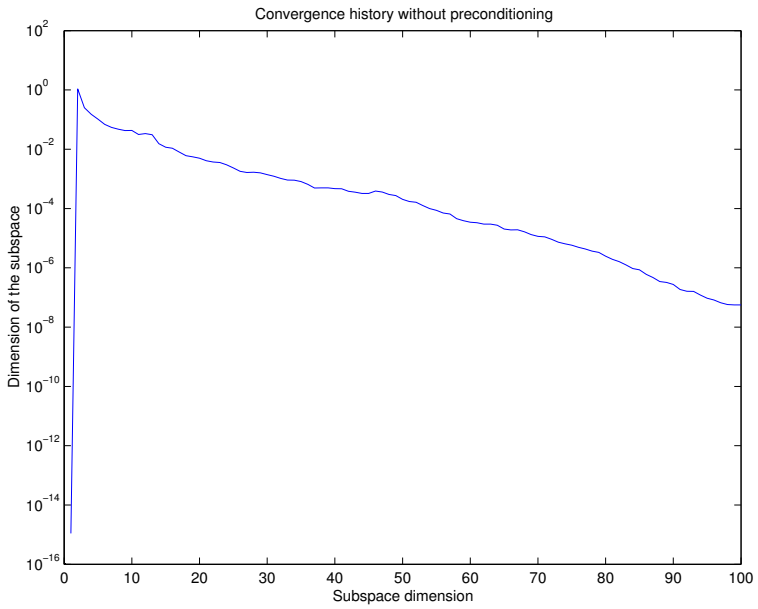

Figure 6.3: Without preconditioning Algorithm 10 converges extremely slow.

6.3 Restarts

The restart strategy was already explained in detail in chapter 5.3. Here we want to demonstrate that this strategy is in fact very effective. We used again a tolerance of 1E-2 for GMRES and performed an update of the LU-decomposition if this accuracy could not be achieved within 5 steps. The tolerance of an eigenpair to be accepted was again 1E-13. A restart was performed after a subspace dimension of 40. The subspace was reduced to the dimension $\max(10, \ell+3)$, where ℓ was the number of the eigenvalue we were currently seeking for. The convergence history with restarts is shown in figure 6.3. With this restart strategy Algorithm 10 needed 15 LU decompositions. This is just one more than without restarts. Overall 382 GMRES steps were performed. This is also slightly more than in the not restarted case. The computation time went down to 14 minutes. Hence, this numerical test shows that the restart strategy for Algorithm 10 is very succesful and allows a good control of the subspace dimension. The only thing to consider is that the dimension of the search subspace must not be smaller than the number of the eigenvalue one is seeking for.

6.4 Eigenvalues in other intervals

In the last test we were seeking for eigenvalues in the interval $(1, 3)$. The test problem has 20 eigenvalues in this interval. The lowest eigenvalue in this interval has the number 11. Hence, the algorithm must start with an 11 dimensional subspace. So we

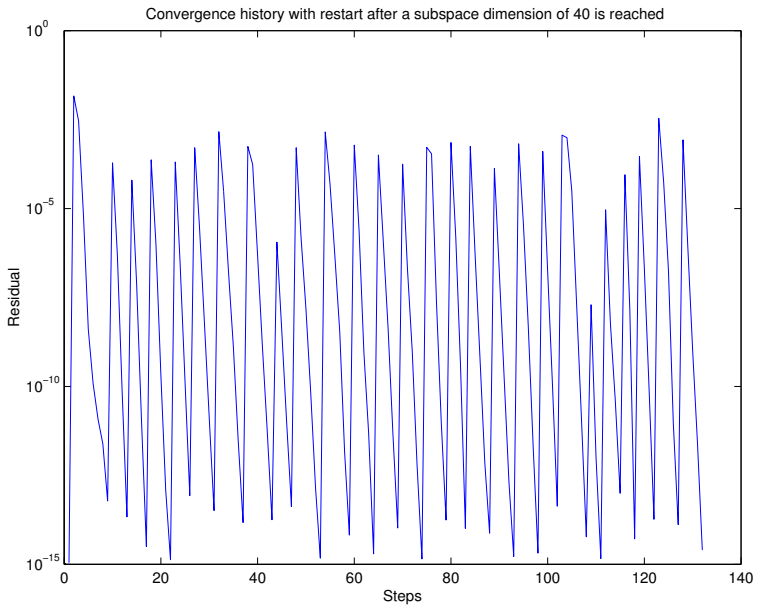

Figure 6.4: Convergence is not significantly slower with restarts.

used as initial search subspace the eleven eigenvectors of the linear problem $Kx = \lambda Mx$ corresponding to the 11 eigenvalues with smallest modulus to maybe have some useful information contained in the subspace. The update of the LU decomposition is again performed if more than 5 GMRES steps are needed and the tolerance for the residual is set to 1E-13. The results are shown in figure 6.4 All eigenvalues are found with a subspace dimension of 107. The algorithm needs 12 LU factorizations and 347 GMRES steps.

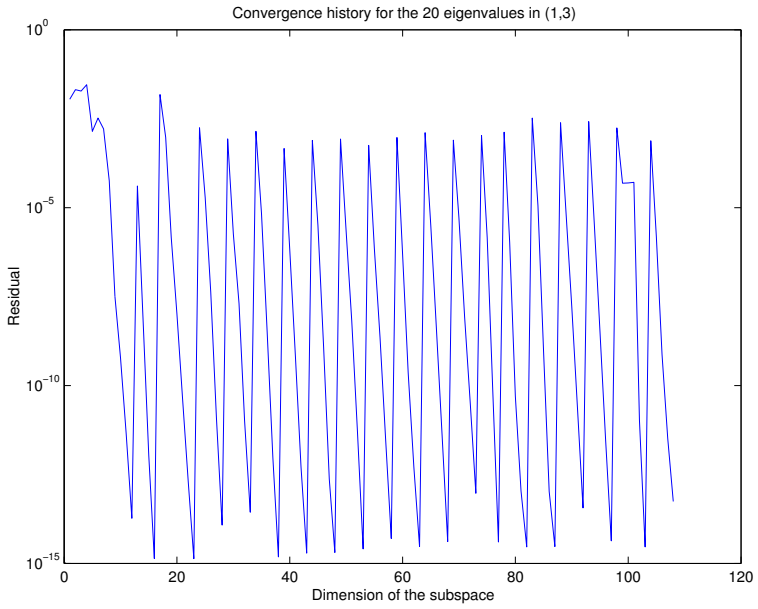

Figure 6.5: The convergence behaviour in the interval $(1, 3)$ is similar to the behaviour in $(0, 1)$.

Chapter 7

Extensions for nonsymmetric problems

For nonsymmetric eigenvalue problems the minmax theory is not valid anymore, since the eigenvalues can be somewhere in the complex plain and need not be ordered. Thus, the guarded iteration for the projected problem in Algorithm 10 can not be applied anymore. But the outer Jacobi-Davidson iteration can also be used for nonsymmetric problems. The analysis of Jacobi-Davidson is also valid for nonsymmetric problems. We just have to find another method for the projected problem to guarantee that the algorithm does not converge to already found eigenvalues in later steps of the iteration.

As example problem we used a problem from vibration analysis of damped systems where the damping is modelled by a viscoelastic constitutive relation (cf. [10],[11]). The nonlinear eigenvalue problem is

$$T(\lambda)x = (\lambda^2 M + K - \frac{1}{1+b\lambda}\Delta K)x = 0 \tag{7.1}$$

The finite element structure used for the tests was the model of an advertising column which has multiple eigenvalues since it is rotational invariant to the z-axis. The matrices K, M and ΔK resulting from the discretisation of the problem are real symmetric 13005×13005 matrices. K and ΔK are positive semidefinite and M is positive definite. For the parameter λ we have $\lambda \in \mathbb{C}$. We are seeking for complex eigenvalues. Therefore the minmax-theory for nonlinear eigenvalue problems can not be applied anymore. The parameter b is a relaxation constant and is set to 0.2E-4 for our numerical tests.

7.1 Comparison with the undamped problem

Since problem (7.1) involves only weak damping, the eigenvectors of the undamped problem

$$Kx = \lambda Mx$$

are already good approximations for the undamped problem. Since the eigenvalues of the linear problem are ordered we use this ordering to separate between eigenvalues of the nonlinear problem. Hence, we use following inner iteration in Algorithm 10 instead of the guarded iteration for symmetric problems.

Algorithm 11 Inner iteration - variant 1

1: Given number j of the wanted eigenvalue and the current subspace V.
2: $\tilde{T}(\lambda) := V^H T(\lambda) V$, $\tilde{K} = V^H K V$, $\tilde{M} := V^H M V$.
3: Solve the projected eigenvalue problem $\tilde{K}S = \tilde{M}S\Lambda$.
4: $x_1 = s_j$, $\lambda_1 = p(x_1)$
5: **for** $k = 1, \ldots$ until convergence **do**
6: $\quad u_{k+1} = \tilde{T}(\lambda_k)^{-1}\tilde{T}'(\lambda_k)$
7: $\quad x_{k+1} = u_{k+1}/\|u_{k+1}\|$
8: $\quad \lambda_{k+1} = p(x_{k+1})$
9: **end for**
10: Return (λ, x) as new approximation for an eigenpair.

In chapter 4.3 we defined the Rayleigh functional only for symmetric eigenvalue problems. But similar to linear eigenvalue problems, a good approximation property of the Rayleigh functional can be observed even for nonsymmetric nonlinear eigenvalue problems. Certainly it must be ensured that the solution of the Rayleigh functional in the interesting region is used. In our example problem, the solution of the Rayleigh functional are roots of a polynomial of degree 3. Since we considered the eigenvalues of the linear problem $\lambda^2 Mx + Kx = 0$ on the negative imaginary axis, we used the complex solution of the Rayleigh functional with negative imaginary part. We applied this strategy to find the first 50 eigenvalues with negative imaginary part. The eigenvalue distribution of the nonlinear problem which was obtained with this variant is shown in figure 7.1. We started the algorithm with the subspace spanned by the eigenvectors for the first ten eigenvalues of the corresponding linear problem to give the algorithm some good start approximations for eigenvectors of the nonlinear problem. We used the same preconditioning strategy as in the symmetric case. Hence, the LU decomposition was updated if GMRES needed more than 5 steps to obtain a relative residual of 1E-3. The accuracy for the residual of an eigenpair to be accepted was set to 1E-3, since the problem only admits eigenvalue approximations with a residual in the region of 1E-4. With this values the algorithm needed 24 LU decompositions and 428 GMRES steps to find all 50 eigenvalues. The computation needed about 50 minutes without

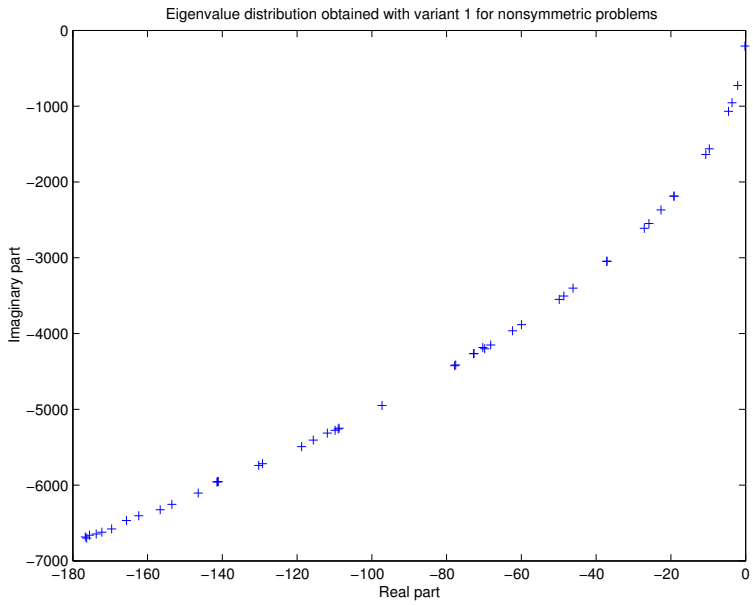

Figure 7.1: The eigenvalue distribution of the nonsymmetric test problem obtained with variant 1

the generation of the start subspace. We used the same test system that was used for the symmetric problem. The convergence history is shown in figure 7.2.

7.2 A new guarding strategy

The inner iteration introduced in the last section works because the problem is only weakly damped and therefore the numbering of the eigenvalues for the linear problem can be used to introduce a numbering for the eigenvalues of the nonlinear problem. This is in general not possible. Therefore we want to introduce a numbering strategy which can be used to find eigenvalues near an arbitrary point in the complex plane.

The numbering of the eigenvalues is based on the idea to compare the distance of the eigenvalues towards a fixed shift in the interesting region of the complex plane. With the shift in figure 7.3 a numbering is introduced for the three interesting eigenvalues based on the distance to the shift. Let σ be the shift and μ_k the current eigenvalue approximation. Then a first order approximation for the eigenvalue problem $T(\lambda)x = 0$ gives

$$T(\lambda)x \approx (T(\mu_k) - hT'(\mu_k))x = 0.$$

In the method of successive linear approximations always the eigenvalue h with smallest modulus is used to proceed with $\mu_{k+1} = \mu_k - h$. Here we want to compare the eigenvalues of the linear problem with the shift σ. Consider that we are interested in the second nearest eigenvalue λ of $T(\cdot)$ according to the shift σ. The value $\mu_k - h$

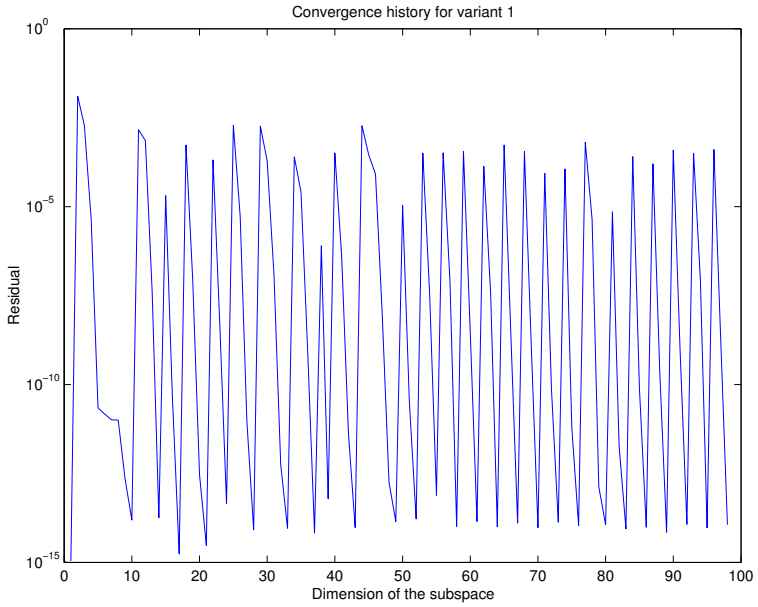

Figure 7.2: The convergence history for variant 1 of the inner iteration

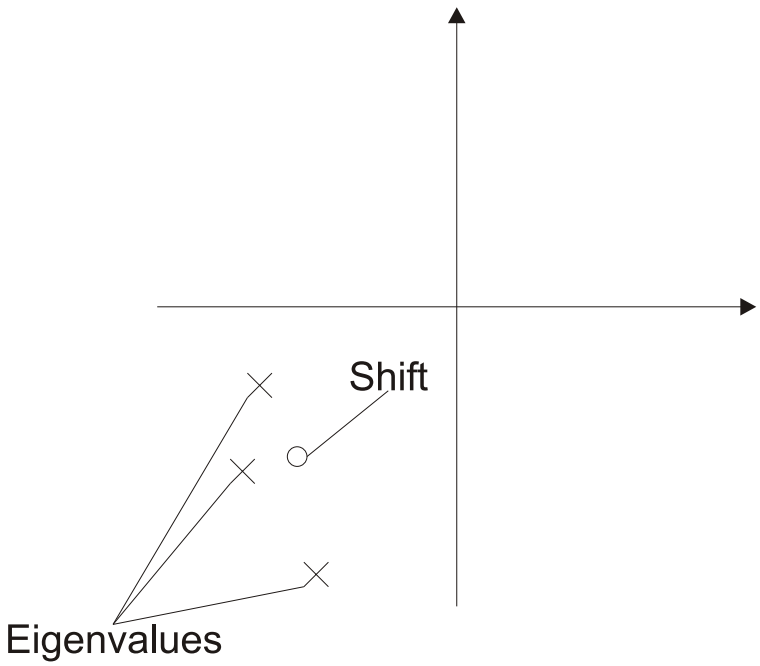

Figure 7.3: A new numbering for complex eigenvalues

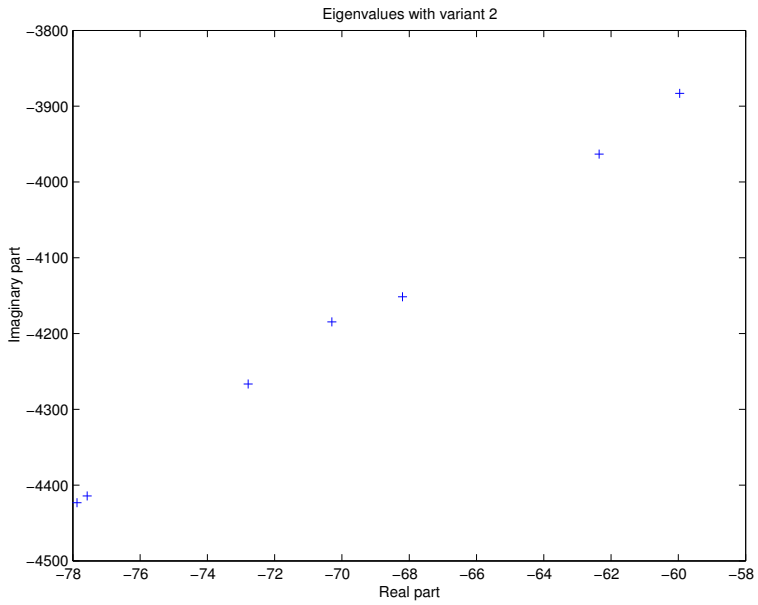

Figure 7.4: The eigenvalue distribution of the nonsymmetric test problem obtained with variant 2

is an approximation for an eigenvalue of the nonlinear problem. Hence, we proceed the iteration with the value $\mu_k - h$ which is second nearest to the shift σ. With this strategy we obtain the following inner iteration.

Algorithm 12 Inner iteration - variant 2

1: Given number j of the wanted eigenvalue, the current subspace V and the shift σ.
2: $\tilde{T}(\lambda) := V^H T(\lambda) V$, $\tilde{K} = V^H K V$, $\tilde{M} := V^H M V$.
3: $\lambda_1 = \sigma$
4: **for** $k = 1, \ldots$ until convergence **do**
5: Solve the eigenvalue problem $\tilde{T}(\lambda_k) S = T'(\tilde{\lambda}_k) S\Theta$.
6: Choose the eigenvalue θ for which the value $|\sigma - (\lambda_k - \theta)|$ is the j-smallest among all eigenvalues
7: $\lambda_{k+1} = \lambda_k - \theta$
8: **end for**

For the numerical test we were seeking for the 10 eigenvalues nearest the shift $\sigma = -60 - 4100i$. As start subspace we used the 10 eigenvectors of the linear problem $Kx = \lambda x$ whose eigenvalues are nearest to the shift σ. The other parameters of the algorithm were set similar to the tests in the last section. We obtained the eigenvalue distribution near the shift shown in figure 7.4. The convergence behaviour is demonstrated in figure 7.5. The algorithm needed 8 LU factorizations and 139 GMRES steps. The overall computation time included the calculation of the start subspace was 22 minutes. The convergence behaviour is not yet as good as in variant 1. Further research is certainly necessary for this guarding strategy. But it already seems to be suitable

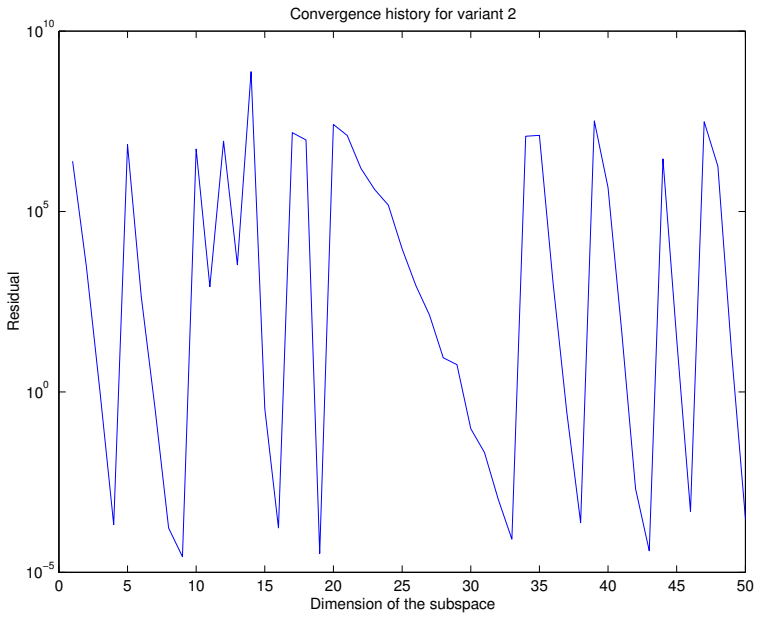

Figure 7.5: The convergence history for 10 eigenvalues with variant 2

for the computation of some eigenvalues near a given shift for arbitrary nonlinear eigenvalue problems.

Chapter 8

Final remarks

The Jacobi-Davidson type approach combined with guarded iteration for symmetric eigenvalue problems is a very capable method to find eigenvalues of large nonlinear eigenvalue problems. The numerical results demonstrate the fast convergence of this method. Even on a normal desktop size computer it was possible to compute the eigenvalues of a nonlinear problem of dimension 36040 in a relatively short time. The restart strategy works very well and allows an efficient control of the dimension of the search subspace. The preconditioning strategy is simple but very effective. Guarding the steps of the GMRES iteration is certainly also for linear eigenvalue problems a good approach to accelerate Jacobi-Davidson by applying a sequence of preconditioners. The only drawback of our new approach is that the dimension of the search subspace always needs to be at least the number of the eigenvalue which shall be extracted. Hence, for eigenvalues with a high number our approach leads to large subspaces which increase computation time. But very often one is interested in smaller eigenvalues and for these eigenvalues our approach leads to very short computation times. The extension to nonsymmetric problems is still subject of research. The approach for weakly undamped problems is already working well. The new guarding method for arbitrary nonlinear eigenvalue problems is also already working. But more tests with different problems are certainly necessary to better understand this guarding strategy.

Acknowledgements

First of all, I want to thank Prof. Voss for the many helpful comments and discussions during the time of the thesis which made the results presented here possible. Further, I want to thank all members of the Department of Mathematics for the support they gave through my whole studies. Being part of the department as student during the last two years was a wonderful experience for me and strongly influenced my fields of interest.

Appendix A

Erklärung

Hiermit versichere ich, daß ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Hamburg, den 20. August 2002

Timo Betcke

Bibliography

[1] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H.A. van der Vorst, editors. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. SIAM, Philadelphia, 200.

[2] E. M. Barston. A minimax principle for nonoverdamped systems. *Int. J. Engng. Sci.*, 12:413–421, 1974.

[3] T. Betcke and H. Voss. A Jacobi-Davidson type projection method for nonlinear eigenvalue problems. Technical Report 47, Department of Mathematics – TU Hamburg-Harburg, Hamburg, Germany, 2002.

[4] Francoise Chatelin. *Eigenvalues of Matrices*. John Wiley & Sons Ltd, 1993.

[5] C. Conca, J. Planchard, and M. Vanninathan. Existence and location of eigenvalues for fluid-solid structures. *Comput. Meth. Appl. Mech. Engrg.*, 77:253–291, 1989.

[6] R. J. Duffin. A minmax theory for overdamped networks. *J. Rat. Mech. Anal.*, 4:221–233, 1955.

[7] J. Eisenfeld and E.H. Rogers. Elementary localization theorems for nonlinear eigenproblems. *J. Math. Anal. Appl.*, 48:325–340, 1974.

[8] K. P. Hadeler. Mehrparametrige und nichtlineare Eigenwertaufgaben. *Arch. Rat. Mech. Anal.*, 27:306–328, 1967.

[9] K. P. Hadeler. Variationsprinzipien bei nichtlinearen Eigenwertaufgaben. *Arch. Rat. Mech. Anal.*, 30:297–307, 1968.

[10] P. Hager and N.E. Wiberg. The Rational Krylov algorithm for nonlinear eigenvalue problems. *Computational Mechanics for the Twenty-First Century*, pages 379 – 402, 2000.

[11] Patrik Hager. *Eigenfrequency Analysis - FE-Adaptivity and a Nonlinear Eigenproblem Algorithm*. PhD thesis, Chalmers University of Technology, Sweden, 2001.

[12] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

[13] V. N. Kublanovskaya. On an approach to the solution of the generalized latent value problem for λ-matrices. *SIAM J. NUMERIC. ANAL.*, 7(4):532–537, 1970.

[14] H. Langer. Über stark gedämpfte Scharen in Hilberträumen. *J. Math. Mech.*, 17:685–705, 1968.

[15] Richard B. Lehoucq, Danny C. Sorensen, and Chao Yang. *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM, Philadelphia, 1998.

[16] A. Neumaier. Residual inverse iteration for the nonlinear eigenvalue problem. *SIAM J. NUMER. ANAL.*, 22(5):914 – 923, October 1985.

[17] J. Planchard. Eigenfrequencies of a tube bundle placed in a confined fluid. *Comput. Meth. Appl. Mech. Engrg.*, 30:75–93, 1982.

[18] E.H. Rogers. A minimax theory for overdamped systems. *Arch. Rat. Mech. Anal.*, 16:89–96, 1964.

[19] E.H. Rogers. Variational properties of nonlinear spectra. *J. Math. Mech.*, 18:479–490, 1968.

[20] Kai Rothe. *Lösungsverfahren für Nichtlineare Matrixeigenwertaufgaben mit Anwendung auf die Ausgleichselementmethode*. PhD thesis, Universität Hamburg, Germany, 1989.

[21] Axel Ruhe. Algorithms for the nonlinear eigenvalue problem. *SIAM J. NUMER. ANAL.*, 10(4):674–689, September 1973.

[22] Axel Ruhe. The Rational Krylov algorithm for generalized eigenvalue problems. In *92 Shanghai International Numerical Algebra and its applications Conference, Fudan University Shanghai, October 26-30, 1992, to appear*, 1993.

[23] Axel Ruhe. Rational Krylov algorithms for nonsymmetric eigenvalue problems, II: Matrix pairs. *Lin. Alg. Appl.*, 197/198:283–296, 1994.

[24] Axel Ruhe. The Rational Krylov for nonlinear eigenvalue problems. In *Talk at the Householder XIII conference, Pontresina, Switzerland*, 1996.

[25] Axel Ruhe. Eigenvalue algorithms with several factorizations - a unified theory yet ? Technical Report 11, Department of Mathematics, Chalmers University of Technology, Göteborg, Sweden, 1998.

[26] Axel Ruhe. Rational Krylov: A practical algorithm for large sparse nonsymmetric matrix pencils. *SIAM J. SCI. COMPUT.*, 19(5):1535–1551, 1998.

[27] Axel Ruhe. A Rational Krylov algorithm for nonlinear matrix eigenvalue problems. In *Zapiski Nauchnyh Seminarov POMI*, volume 268, pages 176–180, 2000.

[28] Y. Saad. *Numerical Methods For Large Eigenvalue Problems*. Manchester University Press, Manchester, 1992.

[29] G. Sleijpen and H.A. van der Vorst. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, chapter 4.7: Jacobi–Davidson Methods, pages 88 – 105. SIAM, Philadelphia, 2000.

[30] G.L.G. Sleijpen, J.G.L. Booten, D.R. Fokkema, and H.A. van der Vorst. Jacobi-Davidson type methods for generalized eigenproblems and polynomial eigenproblems. *BIT*, 36:595–633, 1996.

[31] G.L.G. Sleijpen and H.A. van der Vorst. The Jacobi-Davidson method for eigenvalue problems as an accelerated inexact newton scheme. to appear in IMACS Conference proceedings, February 1995.

[32] G.L.G. Sleijpen and H.A. van der Vorst. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl*, 17:401–425, 1996.

[33] G. W. Stewart. *Introduction to matrix computations*. Academic Press, New York, 1973.

[34] R. E. L. Turner. Variational properties of nonlinear spectra. *J. Math. Anal. Appl.*, 17:151–160, 1967.

[35] R. E. L. Turner. A class of nonlinear eigenvalue problems. *J. Func. Anal.*, 7:297–322, 1968.

[36] H. Voss. A maxmin principle for nonlinear eigenvalue problems with application to a rational spectral eigenvalue problem in fluid-solid vibration. Technical Report 48, Department of Mathematics – TU Hamburg-Harburg, Hamburg, Germany, 2002.

[37] H. Voss and B. Werner. A minimax principle for nonlinear eigenvalue problems with application to nonoverdamped systems. *Math. Meth. Appl. Sci.*, 4:415–424, 1982.

[38] B. Werner. *Das Spektrum von Operatorenscharen mit verallgemeinerten Rayleighquotienten*. PhD thesis, Universität Hamburg, Germany, 1970.

[39] B. Werner. Das Spektrum von Operatorscharen mit verallgemeinerten Rayleighquotienten. *Arch. Rat. Mech. Anal.*, 42:223–238, 1971.