Robust solution methods for nonlinear eigenvalue problems

Thèse

présentée le 29 août 2013

à la Faculté des Sciences de Base
Chaire d'algorithmes numériques et calcul haute performance
Programme doctoral en Mathématiques

École polytechnique fédérale de Lausanne

pour l'obtention du grade de Docteur ès Sciences

par

Cedric Effenberger

Jury:

Prof. Philippe Michel, président du jury
Prof. Daniel Kressner, directeur de thèse
Prof. Peter Benner, rapporteur
Prof. Wolf-Jürgen Beyn, rapporteur
Prof. Marco Picasso, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Lausanne, Suisse
2013

Acknowledgments

The research documented within this thesis was conducted from November 2009 until February 2012 at the Seminar for Applied Mathematics (SAM) of ETH Zurich and from March 2012 until September 2013 at the Mathematics Institute of Computational Science and Engineering (MATHICSE) of EPF Lausanne.

Special thanks are due to my thesis advisor, Prof. Dr. Daniel Kressner, for his enthusiasm and support over the past years. I learned a lot from him, both scientifically and non-scientifically, and it was him who sparked my interest in nonlinear eigenvalue problems.

Furthermore, I thank Prof. Dr. Peter Benner, Prof. Dr. Wolf-Jürgen Beyn, Prof. Dr. Philippe Michel, and Prof. Dr. Marco Picasso for agreeing to serve on my PhD committee.

I would like to express my gratitude to Prof. Dr. Heinrich Voß for his interest in my work, for several fruitful discussions about Jacobi-Davidson algorithms, and the invitation to visit the Hamburg University of Technology in December 2012.

I am also indebted to Prof. Dr. Karl Meerbergen and Prof. Dr. Wim Michiels for our collaboration and the invitation to visit the KU Leuven in March 2013.

Likewise, I am grateful to Prof. Dr. Olaf Steinbach and Dr. Gerhard Unger for our joint work on interpolation-based methods for boundary-element discretizations of PDE eigenvalue problems.

Finally, I thank Michael Steinlechner for providing a code basis for the numerical experiments in Chapter 5 as part of his Bachelor project.

Financial support by the Swiss National Science Foundation under the SNSF research module *Robust numerical methods for solving nonlinear eigenvalue problems* within the SNSF ProDoc *Efficient Numerical Methods for Partial Differential Equations* is thankfully acknowledged.

At the private level, I wish to thank all of my colleagues at SAM and MATHICSE for making my stay at these institutes a great experience. In particular, I appreciated the always pleasant atmosphere created by my office mates, Andrea, Christine, Elke, Jingzhi, Marie, Petar, Shipeng, and Sohrab. Besides, special thanks go to Christine for many interesting conversations about virtually everything, to Roman for his subtle humor and always getting straight to the point, and to Meiyue for sharing a different brain teaser every day.

Last but not least, I am grateful to my parents and my sister for always believing in me, and to Carolin for her support and for scrutinizing the manuscript.

Abstract

In this thesis, we consider matrix eigenvalue problems where the eigenvalue parameter enters the problem in a nonlinear fashion but the eigenvector continues to enter linearly. In particular, we focus on the case where the dependence on the eigenvalue is non-polynomial. Such problems arise in a variety of applications, some of which are portrayed in this work.

Thanks to the linearity in the eigenvectors, the type of nonlinear eigenvalue problems under consideration exhibits a spectral structure which closely resembles that of a linear eigenvalue problem. Nevertheless, there also exist fundamental differences; for instance, a nonlinear eigenvalue problem can have infinitely many eigenvalues, and its eigenvectors may be linearly dependent. These issues render nonlinear eigenvalue problems much harder to solve than linear ones, and even though corresponding algorithms exist, their performance comes nowhere close to that of a modern linear eigensolver.

Recently, minimal invariant pairs have been proposed as a numerically robust means of representing a portion of a nonlinear eigenvalue problem's spectral structure. We compile the existing theory on this topic and develop this concept further. Other major theoretical contributions include a deflation technique for nonlinear eigenvalue problems, a first-order perturbation analysis for simple invariant pairs, as well as the characterization of the generic bifurcations of a real nonlinear eigenvalue problem depending on one real parameter.

Based on these advances in theory, we propose and analyze several new algorithms for the solution of nonlinear eigenvalue problems, which improve on the properties of the existing solvers. Various algorithmic details, such as the efficient solution of the occurring linear systems or the choice of certain parameters, are discussed for the proposed methods. Finally, we apply prototype implementations to a range of selected benchmark problems and comment on the results.

Keywords: nonlinear eigenvalue problem, minimal invariant pair, numerical continuation, polynomial interpolation, deflation, preconditioned eigensolver, Jacobi-Davidson method

Zusammenfassung

Diese Dissertation behandelt Matrixeigenwertprobleme, bei denen der Eigenwertparameter nichtlinear, der Eigenvektor jedoch weiterhin nur linear auftritt. Insbesondere konzentrieren wir uns auf den Fall einer nicht-polynomiellen Abhängigkeit vom Eigenwert. Solche Probleme treten in einer Vielzahl von Anwendungen auf, von denen einige in der Arbeit behandelt werden.

Infolge der linearen Abhängigkeit vom Eigenvektor, ist die spektrale Struktur dieser Art von nichtlinearen Eigenwertproblemen der von linearen Eigenwertproblemen sehr ähnlich. Allerdings gibt es auch fundamentale Unterschiede. So kann ein nichtlineares Eigenwertproblem zum Beispiel unendlich viele Eigenwerte besitzen oder seine Eigenvektoren können linear abhängig sein. Diese Schwierigkeiten verkomplizieren die Lösung nichtlinearer Eigenwertprobleme gegenüber linearen deutlich. Zwar existieren entsprechende Algorithmen, deren Leistungsfähigkeit erreicht jedoch bei weitem nicht die eines modernen linearen Eigenwertlösers.

Die kürzlich vorgeschlagenen minimalen invarianten Paare erlauben eine numerisch stabile Darstellung eines Teils der spektralen Struktur nichtlinearer Eigenwertprobleme. Wir stellen die vorhandene Theorie zu diesem Thema zusammen und entwickeln das Konzept weiter. Ein Deflationsverfahren für nichtlineare Eigenwertprobleme, eine Störungsanalyse erster Ordnung für einfache invariante Paare sowie die Charakterisierung der generischen Bifurkationen eines reellen nichtlinearen Eigenwertproblems mit einem reellen Parameter stellen weitere bedeutende Beiträge zur Theorie dar.

Gestützt auf diese theoretischen Fortschritte, entwickeln und analysieren wir mehrere neue Algorithmen zur Lösung nichtlinearer Eigenwertprobleme, welche die Eigenschaften der bestehenden Löser verbessern. Auf algorithmische Aspekte, wie die effiziente Lösung der auftretenden linearen Systeme oder die Wahl gewisser Parameter, gehen wir dabei detailliert ein. Abschließend erproben wir prototypische Implementationen unserer Verfahren anhand ausgewählter Testprobleme und kommentieren die Ergebnisse.

Schlagwörter: nichtlineares Eigenwertproblem, minimales invariantes Paar, numerische Fortsetzung, Polynominterpolation, Deflation, vorkonditionierter Eigenwertlöser, Jacobi-Davidson Verfahren

Contents

Chapter 1

Introduction

This thesis is concerned with the numerical solution of nonlinear eigenvalue problems of the form

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

with a matrix-valued function $T : \mathcal{D} \to \mathbb{C}^{n \times n}$ defined on some subdomain $\mathcal{D}$ of the complex plane. Any pair (x, λ) satisfying (1.1) is called an eigenpair of T, composed of the (right) eigenvector $x \in \mathbb{C}^n \setminus \{0\}$ and the eigenvalue $\lambda \in \mathcal{D}$. Occasionally, we are also interested in left eigenvectors $y \in \mathbb{C}^n \setminus \{0\}$ satisfying

$$y^{\mathsf{H}} T(\lambda) = 0, \qquad y \neq 0.$$

A triple consisting of an eigenvalue λ and corresponding left and right eigenvectors is called an eigentriple of T. The set of all eigenvalues is commonly known as the spectrum of T and will be denoted by $\operatorname{spec} T$. The complement with respect to $\mathcal{D}$ of the spectrum is called the resolvent set. In this work, we will confine ourselves to regular nonlinear eigenvalue problems, for which the resolvent set is non-empty. One then readily verifies that the mapping $\lambda \mapsto T(\lambda)^{-1}$, called the resolvent of T, is well-defined for all λ in the resolvent set.

The dependence of T on λ is typically nonlinear, hence the name of the problem. On the other hand, the eigenvector x enters the problem only linearly. This linearity in the eigenvector is a crucial feature of the problems considered in this work, and many of the subsequent derivations would not be feasible without this property. Unfortunately, the term *nonlinear eigenvalue problem* is ambiguous in that it also frequently refers to another important class of eigenvalue problems [89, 58, 31], where both the eigenvalue and the eigenvector enter in a nonlinear fashion. The latter type of problems, however, will not be covered in this thesis.

In the special case where the matrix-valued function T is a polynomial in λ, the problem (1.1) is usually called a polynomial eigenvalue problem, or even a quadratic eigenvalue problem if the degree of the polynomial is 2. Likewise, we speak of a rational eigenvalue problem whenever all entries of the matrix $T(\lambda)$ are rational functions in λ. Polynomial and, in particular, quadratic eigenvalue problems have received a lot of attention in the literature; see, e.g., [48, 127, 93]. Rational eigenvalue problems, in turn, can be traced back to polynomial eigenvalue problems through multiplication by the common denominator. Considerably less work has been done in the direction of more general nonlinearities which do not fall into one of the aforementioned categories. These general nonlinear eigenvalue problems will be the focus of the present thesis. In fact, with the exception of Chapter 7, the theory and algorithms developed in this work will only assume that the matrix-valued function T depends holomorphically on λ. Apart from this requirement, no further restrictions are imposed on the nonlinearity. To emphasize this fact, we will sometimes also speak about the general holomorphic eigenvalue problem (1.1).

1.1 Linear vs. nonlinear eigenvalue problems

For the specific choice $T(\lambda) = \lambda I - A$, the nonlinear eigenvalue problem (1.1) reduces to the standard linear eigenvalue problem for the matrix $A \in \mathbb{C}^{n\times n}$. In this sense, nonlinear eigenvalue problems can be regarded as a generalization of the standard eigenvalue problem, and it will be interesting to see what properties of the standard eigenvalue problem carry over to the nonlinear setting.

The numerical solution of linear eigenvalue problems has been studied extensively over the last decades and can now be considered a mature topic in numerical analysis. It is well-known that any matrix $A \in \mathbb{C}^{n\times n}$ has at most n distinct eigenvalues, which are the roots of the characteristic polynomial $p_A(\lambda) = \det(\lambda I - A)$ associated with A. The multiplicity of an eigenvalue λ as a root of p_A is called the algebraic multiplicity of λ, denoted by $\operatorname{alg}_A \lambda$. When we take these multiplicities into account, then every $A \in \mathbb{C}^{n\times n}$ has exactly n eigenvalues. The eigenvectors belonging to the same eigenvalue λ (plus the zero vector) constitute a subspace of $\mathbb{C}^n$—the eigenspace associated with λ. The dimension of this eigenspace is called the geometric multiplicity of the eigenvalue λ and can be shown to be either equal to or less than the algebraic multiplicity of λ. In case it is less, we can complement the eigenvectors with generalized eigenvectors by forming Jordan chains until the dimension of the generalized eigenspace spanned by the eigenvectors and generalized eigenvectors equals the algebraic multiplicity of λ. In the remainder of this work, we will write *(generalized) eigenvectors* as a shorthand for *eigenvectors and generalized eigenvectors*. The (generalized) eigenvectors belonging to different eigenvalues of the matrix A are linearly independent from each other. This is an extremely important property because it permits us to compile a basis of $\mathbb{C}^n$ comprising only (generalized) eigenvectors of A. The representation of the matrix A with respect to such a basis exhibits a very special structure commonly known as the Jordan canonical form of A.

Some of the concepts introduced in the previous paragraph can be extended in a straightforward manner to general holomorphic eigenvalue problems; see also, e.g., [98]. Clearly, the eigenvalues of the nonlinear eigenvalue problem (1.1) are the roots of the characteristic equation

$$\det T(\lambda) = 0. \tag{1.2}$$

Consequently, we call the multiplicity of an eigenvalue λ as a root of (1.2) the algebraic multiplicity of λ, denoted as $\operatorname{alg}_T \lambda$. It has been shown in [8, 126] that any regular holomorphic eigenvalue problem has a discrete spectrum consisting of eigenvalues with finite algebraic multiplicities. However, the number of eigenvalues is, in general, not related to the problem size. In particular, there may be infinitely many eigenvalues, as the example

$$T : \mathbb{C} \to \mathbb{C}, \qquad T(\lambda) = \sin(\lambda) \tag{1.3}$$

indicates. As in the linear case, the eigenvectors belonging to the same eigenvalue λ constitute a subspace of $\mathbb{C}^n$ in conjunction with the zero vector. We will once more refer to this subspace as the eigenspace associated with λ and to its dimension as the geometric multiplicity of λ. It remains true that the geometric multiplicity of an eigenvalue is bounded by the algebraic multiplicity. When there is a deficit, we may again form Jordan chains to find additional generalized eigenvectors. However, nonlinear eigenvalue problems suffer from a severe loss of linear independence among their (generalized) eigenvectors. More specifically, this loss is twofold. On the one hand, generalized eigenvectors need no longer be linearly independent from the eigenvectors belonging to the same eigenvalue. In fact, all generalized eigenvectors may be zero. On the other hand, (generalized) eigenvectors belonging to different eigenvalues are not guaranteed to be linearly independent anymore. This fact is obvious in situations with infinitely many eigenvalues but may occur as well for finite spectra. For instance, the quadratic eigenvalue problem

$$\left(\lambda^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \lambda \begin{bmatrix} -1 & -6 \\ 2 & -9 \end{bmatrix} + \begin{bmatrix} 0 & 12 \\ -2 & 14 \end{bmatrix}\right) x = 0, \qquad x \neq 0 \tag{1.4}$$

taken from [34, Example 4.3] has the four eigenvalues 1, 2, 3, and 4. In this example, the eigenvalues 3 and 4 share the same eigenvector $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$.

For linear eigenvalue problems, linear independence of the (generalized) eigenvectors plays an essential role in the reliable computation of several eigenpairs. Eigenvectors which have already been determined are deflated from the problem by restricting the computation to their orthogonal complement, ensuring that they will not be found again. For nonlinear eigenvalue problems, such an approach would be improper as it bears the risk of missing eigenvalues. For example, when admitting only one copy of the eigenvector $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ in the solution of the quadratic eigenvalue problem (1.4), only one of the eigenvalues 3 and 4 can be found. As a result, most existing solution algorithms for general nonlinear eigenvalue problems (see [97, 112] as well as the next section) have difficulties with computing several eigenpairs in a robust and dependable manner.

Unlike their linear counterpart, nonlinear eigenvalue problems can be quite challenging already at small scales. Whereas the QR algorithm routinely computes all eigenvalues and eigenvectors of a small linear problem, a similar algorithm does not exist in the general nonlinear case. Part of the reason is that even a nonlinear eigenvalue problem of size 1×1, such as (1.3), can have infinitely many eigenvalues. Hence, one cannot expect to compute the entire spectrum. The lack of a black-box solver for small-scale nonlinear eigenvalue problems also has an adverse effect on large-scale nonlinear eigenvalue solvers because these frequently involve the solution of (a sequence of) smaller subproblems.

The additional level of complexity incurred by the nonlinear dependence on the eigenvalue parameter manifests itself as well in the absence of widely avaliable, dedicated software. While most popular eigensolver packages, such as SLEPc [59], ARPACK [87], or Anasazi [11], supply routines for many different flavors of linear eigenvalue problems, their support for nonlinear eigenvalue problems is either non-existent or confined to polynomial or even quadratic eigenvalue problems. As far as the author is aware, no software has been released for general nonlinear eigenvalue problems beyond mere research code.

1.2 Existing solution algorithms

Despite the shortage of software implementations, a range of solution techniques for general holomorphic eigenvalue problems can be found in the literature. We will briefly review them here.

1.2.1 Newton-based methods. A broad class of nonlinear eigensolvers are based on Newton iterations. Specifically, there are two variants. Either Newton's method is utilized to find roots of the (scalar) characteristic equation (1.2) or it is applied directly to the (vectorial) nonlinear eigenvalue equation (1.1). For the first variant, it has been proposed in [84] to compute the determinant $\det T(\lambda)$ together with the Newton correction for λ by performing a rank-revealing LQ (lower triangular-orthogonal) decomposition of $T(\lambda)$ at every iteration. This procedure, however, is computationally feasible only for small to medium-sized problems.

In the second variant, the eigenvalue equation needs to be complemented with an appropriate normalization condition $v^{\mathsf{H}}x = 1$, $v \in \mathbb{C}^n$ on the eigenvector x to enforce $x \neq 0$. The application of Newton's method to the combined system then leads to the *nonlinear inverse iteration* [106, 4]

$$\tilde{x}_{j+1} = T(\lambda_j)^{-1}\dot{T}(\lambda_j)x_j, \quad \lambda_{j+1} = \lambda_j - \frac{v^{\mathsf{H}}x_j}{v^{\mathsf{H}}\tilde{x}_{j+1}}, \quad x_{j+1} = \frac{\tilde{x}_{j+1}}{\|\tilde{x}_{j+1}\|}. \tag{1.5}$$

Here and throughout this work, $\dot{T}$ signifies the derivative of the matrix-valued function T with respect to the parameter λ. The scaling of the eigenvector iterates is performed only to avoid possible over- or underflows. Any convenient norm $\|\cdot\|$ can be used for this purpose. Every iteration of the nonlinear inverse iteration (1.5) requires the solution of a linear system of equations. However, since λ_j varies with the iteration number j, the system matrix is different each time. This may be disadvantageous in conjunction with direct solvers because it precludes the reuse of existing factorizations. Unfortunately, replacing λ_j by a fixed shift $\sigma \in \mathbb{C}$ leads to erratic convergence of the method, as observed, e.g., in [134].

A second-order modification of the nonlinear inverse iteration proposed in [101] yields the *residual inverse iteration*

$$x_{j+1} = x_j - T(\lambda_j)^{-1}T(\lambda_{j+1})x_j, \quad \text{where} \quad v^{\mathsf{H}}T(\lambda_{j+1})x_j = 0. \tag{1.6}$$

This modified iteration has the advantage that $T(\sigma)^{-1}$ with a fixed shift σ can be substituted for $T(\lambda_j)^{-1}$ without destroying convergence to the correct solution. Again, the eigenvector iterates should be normalized, at least periodically, to prevent over- or underflows.

As variants of Newton's method, both the nonlinear inverse iteration and the residual inverse iteration exhibit a local and asymptotically quadratic convergence towards simple eigenvalues. In contrast, the simplified residual inverse iteration using a fixed shift σ converges only linearly. Its convergence rate is proportional to the distance between the shift and the closest eigenvalue [101]. Although the slower convergence tends to increase the number of iterations needed, this may be more than offset by the diminished cost of every individual iteration.

All of the above methods are single-vector iterations; i.e., only one eigenvector approximation is stored at a time. This approximation is iteratively corrected until convergence towards a true eigenvector. Whenever a new approximation has been computed, the previous one is discarded. Thus, memory consumption is kept to a necessary minimum at the expense of loosing some information in every step. If memory is not a concern, it may be beneficial to retain the previous approximations and capitalize on this information. The next iterate is then extracted from the subspace spanned by all previous eigenvector approximations via a Rayleigh-Ritz procedure. This idea has led to the *nonlinear Arnoldi algorithm* [134]. The use of subspace acceleration also opens up the possibility for inexact solves of the (simplified) Newton correction equation, eventually giving rise to the *nonlinear Jacobi-Davidson method* [18, 137].

Iterative correction schemes are not limited to the eigenvector. It is similarly possible to devise algorithms based on iterative correction of an approximation to the eigenvalue. For instance, linearizing the matrix-valued function T in the eigenvalue equation (1.1) suggests the update

$$\lambda_{j+1} = \lambda_j - \mu_j, \quad \text{where} \quad T(\lambda_j)x_j = \mu_j \dot{T}(\lambda_j)x_j, \quad x_j \neq 0.$$

This iteration is called the *method of successive linear problems* [112] because we have to solve a (generalized) linear eigenvalue problem in every step—as opposed to a linear system of equations with the eigenvector-based algorithms. If μ_j is chosen as the eigenvalue with smallest modulus of the pair $\big(T(\lambda_j), \dot{T}(\lambda_j)\big)$, the iteration converges locally quadratically.

All algorithms treated so far are directed towards computing one eigenpair only. Although several runs of a method could, in principle, return several eigenpairs, no precautions are taken to prevent the method from repeatedly converging to the same eigenpair. This reconvergence tremendously complicates the robust computation of clustered eigenvalues. In the extreme case of a multiple and defective eigenvalue, it is even impossible to discover the generalized eigenvectors. Block versions of the aforementioned algorithms provide a way around this obstacle in a similar manner as subspace iteration for linear eigenvalue problems. In [83], a *block Newton method* has been proposed as a block analog of the nonlinear inverse iteration. Block methods avoid the difficulties with reconvergence by computing all eigenvalues in a cluster simultaneously. Unfortunately, the local convergence of these methods seems to be more restricted than that of their single-vector counterparts. Furthermore, the block size, i.e., the number of eigenvalues in the cluster, needs to be specified in advance.

1.2.2 Methods based on contour integration. A rather different solution approach for nonlinear eigenvalue problems is founded on the observation [49] that the resolvent of T is a finitely meromorphic function, whose poles are exactly the eigenvalues of T. Consequently, the resolvent possesses a Laurent series expansion in a suitable punctured neighborhood of any eigenvalue $\mu \in \operatorname{spec} T$. According to Keldysh's theorem, which has been proved in [75] for polynomial eigenvalue problems and extended to the general holomorphic setting in [49], the principal part of this Laurent series can be expressed in terms of the (generalized) left and right eigenvectors associated with μ. It is therefore appealing to extract this spectral information from the resolvent by means of contour integration. More precisely, let $\mathcal{C}$ be a contour in the domain of definition of T not passing through any eigenvalues and consider, for $k = 0, 1, 2, \dots$, the integrals

$$A_k = \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{C}} \xi^k T(\xi)^{-1} \, \mathrm{d}\xi. \tag{1.7}$$

If the (generalized) eigenvectors belonging to the eigenvalues of T enclosed by the contour $\mathcal{C}$ are linearly independent, then the rank of A_0 equals the cumulated algebraic multiplicities of the enclosed eigenvalues. Performing an economy-size singular value decomposition $A_0 = V\Sigma W^{\mathsf{H}}$, the eigenvalues of the matrix pencil $V^{\mathsf{H}} A_1 W - \lambda\Sigma$ then match the eigenvalues of T enclosed by the contour $\mathcal{C}$. If the (generalized) eigenvectors belonging to the enclosed eigenvalues fail to be linearly independent, the above procedure has to be carried out with A_0 and A_1 replaced by block Hankel matrices of the form [19, 7]

$$\mathcal{A}_0 = \begin{bmatrix} A_0 & \cdots & A_{\ell-1} \\ \vdots & \ddots & \vdots \\ A_{\ell-1} & \cdots & A_{2\ell-2} \end{bmatrix}, \quad \mathcal{A}_1 = \begin{bmatrix} A_1 & \cdots & A_{\ell} \\ \vdots & \ddots & \vdots \\ A_{\ell} & \cdots & A_{2\ell-1} \end{bmatrix}$$

for some suitably chosen integer ℓ.

In a practical routine, the contour integral is approximated by the trapezoidal rule, which, in this case, leads to an exponential decline of the quadrature error with the number of quadrature nodes [19]. Still, numerical experiments indicate that, under certain circumstances, the amount of quadrature nodes needed can be quite high [128]. Moreover, computing the full inverse $T(\xi)^{-1}$ in the contour integral (1.7) is intractable for larger problems. Practical implementations therefore work with $A_k\hat{W}$ [19] or $\hat{V}^{\mathsf{H}} A_k \hat{W}$ [7] instead of A_k, where $\hat{V}, \hat{W} \in \mathbb{C}^{n\times\hat{r}}$ are random matrices with a sufficiently large number of columns, $\hat{r}$. Although this replacement reduces the computational effort to essentially $\hat{r}$ linear solves per quadrature node, contour integral-based methods still tend to be rather computationally demanding. On the other hand, these methods are easily parallelizable because the computations at different quadrature nodes are completely independent from each other [2]. In addition, contour integral-based methods have the unique advantage that they can guarantee the discovery of *all* eigenvalues within a prescribed region of the complex plane.

1.2.3 Infinite Arnoldi methods. It is well-known, see, e.g., [48, 93], that any polynomial eigenvalue problem can be reformulated as an equivalent linear eigenvalue problem of larger dimension. This conversion is commonly referred to as *linearization*. Recently, a similar technique has been proposed [69] also for general holomorphic eigenvalue problems, resulting in linear operator eigenvalue problems on an infinite-dimensional function space. When applying Arnoldi's method in this infinite-dimensional setting, the Arnoldi vectors are functions and therefore, normally, cannot be stored (exactly) in a computer's memory. To overcome this obstacle, the *infinite Arnoldi method* [69] chooses a specific starting vector, which ensures that the Arnoldi vectors can be represented by finitely many coefficients as linear combinations of polynomials and exponentials. Later, several extensions to the infinite Arnoldi algorithm have been published, including a modification to enable locking and restarts [67] as well as a generalization to a certain class of infinite-dimensional nonlinear eigenvalue problems [66].

1.2.4 Interpolation-based methods. Since polynomial eigenvalue problems can be conveniently solved via linearization [48, 93], it is tempting to (locally) approximate eigenvalue problems with more general nonlinearities by polynomial ones. In applications, such polynomial approximations are often gained through truncated Taylor expansion; see [78, 72, 79, 104, 124] for examples arising from boundary integral formulations of partial differential (PDE) eigenvalue problems. However, one can as well employ other, more sophisticated approximation schemes. In [25], a polynomial interpolation approach combined with principal component analysis, similarly as in the discrete empirical interpolation method [29], is proposed for Helmholtz eigenvalue problems with transparent-influx boundary conditions. In [130], a Hermite interpolation strategy is pursued where the interpolation nodes are allowed to be selected incrementally, e.g., guided by the Ritz values obtained in previous runs. Finally, we will develop an algorithm utilizing polynomial interpolation in Chebyshev nodes of either the first or the second kind in Chapter 5.

1.2.5 Methods based on Rayleigh functionals. Hermitian linear eigenvalue problems have the feature that their eigenvectors are the stationary points of the associated Rayleigh quotient. Beyond that, their eigenvalues are real and can be characterized by three variational principles: Rayleigh's principle, Poincaré's principle and the principle of Courant, Fischer, and Weyl. These variational characterizations have far-reaching consequences. In particular, they imply that the smallest eigenvalue is identical to the minimum of the Rayleigh quotient, paving the way for powerful eigensolvers based on Rayleigh quotient minimization, such as the locally optimal (block) preconditioned conjugate gradient method, LO(B)PCG [80].

For some nonlinear eigenvalue problems, analogous constructions are possible. However, in the nonlinear case, it is insufficient to require that $T(\lambda)$ be Hermitian for any λ, as is already evident from the fact that this condition alone does not guarantee the spectrum to be real. In fact, we additionally have to assume the existence of a so-called Rayleigh functional. Rayleigh functionals have been introduced in [38] for quadratic eigenvalue problems and generalized to broader classes of Hermitian nonlinear eigenvalue problems in [111, 55, 138]. In the special case where $T(\lambda) = \lambda I - A$ represents a linear eigenvalue problem for some Hermitian matrix $A \in \mathbb{C}^{n \times n}$, the notion of a Rayleigh functional coincides with the ordinary Rayleigh quotient associated with A.

The existence of a Rayleigh functional ρ allows the three variational principles from the linear, Hermitian case to be generalized to the nonlinear setting; see [56] for Rayleigh's principle, [139] for Poincaré's principle, and [132] for the principle of Courant, Fischer, and Weyl. The variational characterization of eigenvalues again has fundamental consequences. Similarly as in the linear case, it enables eigensolvers based on Rayleigh functional minimization. Variants of the *preconditioned inverse iteration*, the *preconditioned steepest descent method*, as well as the *locally optimal preconditioned conjugate gradient method* for a special class of nonlinear eigenvalue problems have been devised and analyzed in [122]. However, it should be mentioned that unlike for linear, Hermitian eigenvalue problems, there exist certain degenerate situations where the smallest eigenvalue and the minimum of the Rayleigh functional do not match.

As another important effect, the variational principles facilitate the development of globally convergent nonlinear eigensolvers. For instance, the *safeguarded iteration* defined by

$$\lambda_{j+1} = \rho(x_j), \quad \text{where} \quad x_j \in \operatorname*{arg\,max}_{x^{\mathsf{H}}x=1} x^{\mathsf{H}} T(\lambda_j) x,$$

converges globally to the smallest eigenvalue [140, 131]. In the absence of a Rayleigh functional, global convergence poses a challenge to most existing solvers for nonlinear eigenvalue problems.

Finally, the variational principles provide a natural numbering for the eigenvalues and limit their quantity to the problem dimension. This numbering can be exploited to verify whether all eigenvalues in a given interval have been computed, rendering missing eigenvalues a non-issue. Such a feature is much sought-after in many applications but can be delivered only by very few algorithms.

1.3 Contributions of this thesis

Chapter 2. We identify a number of sources for nonlinear eigenvalue problems from various fields of science and technology and give corresponding references to the literature. The focus is on such problems where the dependence on the eigenvalue is non-polynomial. A cross section of four applications is examined more closely. These applications later form the basis for the numerical experiments in the subsequent chapters.

Chapter 3. Minimal invariant pairs constitute an indispensable tool for the numerical representation of several eigenpairs for a nonlinear eigenvalue problem. Chapter 3 gives a new and more intuitive justification of the concept, compiles the existing theory and makes various new contributions.

First, we establish an improved characterization of minimality, which eliminates potential numerical difficulties arising from the use of the monomial basis in the original formulation. The definition of invariance is also slightly generalized by employing contour integration.

Second, we formalize the link between a minimal invariant pair and the Jordan structure of the underlying nonlinear eigenvalue problem in a mathematically rigorous way. This work generalizes the theory of Jordan pairs and constitutes the fundament for the usage of minimal invariant pairs in theoretical developments as well as numerical algorithms.

Third, we examine composite pairs and relate their properties to those of their constituents. This analysis opens up a connection between composite pairs and the derivatives of certain residuals, which can be exploited to redesign the user-interface of a broad class of nonlinear eigensolvers.

Finally, nesting of minimal invariant pairs is considered. We advance an existing technique for extracting a minimal invariant pair from a non-minimal one. This is complemented by a method to embed a given minimal invariant pair into a larger pair which is complete in a certain sense. We show by means of an example from the theory of nonlinear Sylvester operators that a combination of these techniques can lead to very elegant and intuitive proofs.

Chapter 4. Motivated by an example taken from wave propagation in periodic media, we investigate nonlinear eigenvalue problems depending on one real design parameter. The goal is to compute and track several eigenvalues of interest as this parameter varies.

Based on the concept of minimal invariant pairs from Chapter 3, a theoretically sound and reliable numerical continuation procedure is developed. To this end, the minimal invariant pair representing the eigenvalues of interest is characterized as a zero of an appropriate nonlinear function. This zero is then continued by means of a standard pseudo-arclength continuation algorithm.

Particular attention is paid to the situation when the Jacobian of the nonlinear function becomes singular. For the real case, it is proven that, generically, such a singularity occurs only if a real eigenvalue represented in the continued minimal invariant pair collides with another real eigenvalue which is not represented. It is shown how this situation can be handled numerically by a suitable expansion of the minimal invariant pair. Finally, the viability of the constructed continuation procedure is illustrated by two numerical examples related to the stability analysis of time-delay systems.

Chapter 5. Many existing nonlinear eigensolvers rely on frequent formation of the residual. Hence, these methods are not well suited for problems where the evaluation of the residual is very costly. Examples of such problems include boundary-element discretizations of operator eigenvalue problems, for which the computation of the residual involves a singular integral with a nonlocal kernel function for every element in the underlying boundary mesh.

As an alternative, we propose to approximate the nonlinear eigenvalue problem at hand by a polynomial one. Our approach is intended for situations where the eigenvalues of interest are located on the real line or, more generally, on a prescribed curve in the complex plane, enabling the use of interpolation in Chebyshev nodes. For stability reasons, the resulting polynomial interpolant is represented in the Chebyshev basis. We solve the polynomial eigenvalue problem by applying Krylov subspace methods to a suitable linearization and show how the arising linear systems can be solved efficiently.

To investigate the error incurred by the polynomial approximation, a first-order perturbation analysis for nonlinear eigenvalue problems is performed. Combined with an approximation result for Chebyshev interpolation, this shows exponential convergence of the obtained eigenvalue approximations with respect to the degree of the approximating polynomial. Furthermore, we discuss the numerically stable extraction of a minimal invariant pair for the polynomial eigenvalue problem from a minimal invariant pair of the linearization. Finally, the method is applied to two synthetic benchmark problems drawn from the field of boundary-element methods to demonstrate its viability.

Chapter 6. Newton-based methods are well-established techniques for the solution of nonlinear eigenvalue problems. If a larger portion of the spectrum is sought, however, their tendency to reconverge to previously determined eigenpairs is a hindrance. To overcome this limitation, we propose and analyze a deflation strategy for nonlinear eigenvalue problems, based on the concept of minimal invariant pairs from Chapter 3.

In the second half of the chapter, this deflation strategy is incorporated into a Jacobi-Davidson-type framework. Various algorithmic details of the resulting method, such as the efficient solution of the correction equation, the solution of the projected problems, and restarting are discussed. Finally, the efficiency of the approach is demonstrated by a sequence of numerical examples.

Chapter 7. As is well-known, the eigenvalues of a Hermitian matrix are the stationary values of the associated Rayleigh quotient. In particular, this fact enables eigensolvers based on Rayleigh quotient minimization. For a specific class of nonlinear eigenvalue problems, similar constructions are possible. Here, the role of the Rayleigh quotient is played by a so-called Rayleigh functional.

For this special class of nonlinear eigenproblems, we consider a preconditioned version of the residual inverse iteration. A convergence analysis of this method is performed, extending an existing analysis of the residual inverse iteration. As an essential ingredient for the convergence analysis, we prove a perturbation result for the Rayleigh functional in the vicinity of an eigenvector, which generalizes an existing result in that direction. The chapter is concluded with a numerical experiment, which demonstrates that the proposed method has the potential to yield mesh-independent convergence for a sequence of discretizations of an operator eigenvalue problem.

Chapter 2

Applications of nonlinear eigenvalue problems

Nonlinear eigenvalue problems of the form (1.1) arise in a multitude of diverse applications from science and technology, such as acoustic field simulations [96], computational quantum chemistry [136], structural dynamics [118], electromagnetic modeling of particle accelerators [88], vibrations of fluid-solid structures [133, 129], or stability analysis of time-delay systems [99, 65] (see also Section 2.1). In finite-element discretizations of operator eigenvalue problems, nonlinearities are often caused by λ-dependent boundary conditions [25, 98] (see also Section 2.4), λ-dependent material parameters [43, 100] (see also Section 2.2), or the use of special basis functions [17, 73]. Boundary-element discretizations, on the other hand, can lead to nonlinear eigenvalue problems even if the underlying operator eigenvalue problem is linear [124]; see also Section 2.3. In the present chapter, we will examine a few of these applications more closely. For a more comprehensive overview of sources for nonlinear eigenvalue problems of type (1.1), see [97] or [15].

2.1 Stability analysis of time-delay systems

Time-delay systems of retarded type are dynamical systems whose evolution does not only depend on the current state of the system but also on its history. If this dependence is concentrated on one or more discrete points in the past (instead of distributed over one or more time intervals), such delay systems can be modeled by a delay differential equation of the form

$$\dot{u}(t) = A_0 u(t) + A_1 u(t-\tau_1) + \cdots + A_k u(t-\tau_k) \tag{2.1}$$

with the discrete delays $0 < \tau_1 < \cdots < \tau_k$ and square matrices $A_0, \dots, A_k \in \mathbb{R}^{n \times n}$. For further details about time-delay systems as well as delay differential equations and their corresponding initial value problems, we refer to [99].

The ansatz $u(t) = \mathrm{e}^{\lambda t} x$ with $x \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ constitutes a non-trivial solution of the delay differential equation (2.1) if and only if (x, λ) is an eigenpair of the associated delay eigenvalue problem

$$\bigl(\lambda I - A_0 - \mathrm{e}^{-\tau_1 \lambda} A_1 - \cdots - \mathrm{e}^{-\tau_k \lambda} A_k\bigr) x = 0. \tag{2.2}$$

This connection suggests that the behavior of the time-delay system described by the delay differential equation (2.1) can be analyzed by investigating the spectral properties of the delay eigenvalue problem (2.2). Indeed, [99, Proposition 1.6] shows that the null solution of the delay differential equation (2.1) is asymptotically stable if and only if all eigenvalues of the delay eigenvalue problem (2.2) are located in the open left half-plane. Consequently, the eigenvalues of interest for delay eigenvalue problems are the ones with largest real part.

Clearly, the matrix-valued function behind the delay eigenvalue problem (2.2) is holomorphic in λ but neither polynomial nor rational. Thus, delay eigenvalue problems constitute an ideal prototype application for the developments in this work.

A few qualitative features are known about the spectrum of a delay eigenvalue problem. First, there are only finitely many eigenvalues to the right of any vertical line in the complex plane [99, Corollary 1.9]. Second, every eigenvalue λ satisfies [99, Proposition 1.10]

$$|\lambda| \leq \|A_0\| + \|A_1\|\mathrm{e}^{-\tau_1\lambda} + \cdots + \|A_k\|\mathrm{e}^{-\tau_k\lambda}.$$

This relation can be employed to construct an envelope curve, which encloses the spectrum.

2.2 Photonic crystals

Photonic crystals are periodic arrangements of dielectric materials with different relative permittivities ϵ. If the periodicity interval is sufficiently small compared to the wavelength of a photon, these structures have the potential to influence the photon's propagation in a similar manner as semiconductors do with electrons. Specifically, this is achieved by tuning the material parameters of the constituents in such a way that wave propagation is inhibited in certain directions and at certain frequencies.

The waves allowed to propagate within a photonic crystal are determined by a PDE eigenvalue problem derived from the macroscopic Maxwell equations; see [70]. For the idealized case that the photonic crystal is composed of lossless materials with frequency-independent permittivities, numerical techniques for solving these PDE eigenvalue problems are well established [44, 9, 37, 116]. The more realistic case of lossy materials whose permittivity may depend on the frequency of the propagating wave gives rise to nonlinear eigenvalue problems and has been investigated numerically only rather recently [113, 100, 61, 92, 43].

Here, we will restrict the discussion to two-dimensional photonic crystals; i.e., the permittivity ϵ is periodic with respect to two spatial coordinates, say x_1 and x_2, and constant with respect to the third, x_3. By decomposing the electromagnetic wave (E, H) into transverse-electric (TE) polarized modes $(E_1, E_2, 0, 0, 0, H_3)$ and transverse-magnetic (TM) polarized modes $(0, 0, E_3, H_1, H_2, 0)$, the full three-dimensional Maxwell equations are reduced to scalar, two-dimensional Helmholtz equations for H_3 and E_3, respectively. For the sake of simplicity, we will focus on TM polarized waves, but the overall constructions are applicable to the TE case as well.

For a TM polarized wave, the third component of the electric field satisfies the

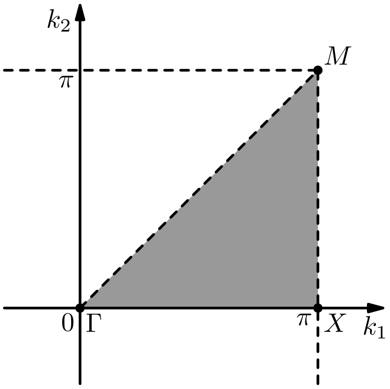

Figure 2.1: The irreducible Brillouin zone with the symmetry points Γ, X, and M.

equation

$$-\Delta E_3(\vec{x}) - \left(\frac{\omega}{c}\right)^2 \epsilon(\vec{x},\omega) E_3(\vec{x}) = 0, \quad \vec{x} = (x_1, x_2), \tag{2.3}$$

where ω is the time frequency and c denotes the speed of light in vacuum. By scaling the spatial coordinates, we may assume that $\epsilon(\cdot,\omega)$ is periodic with respect to the lattice $\mathcal{L} = \mathbb{Z}^2$ with the unit cell $\Omega = (0,1]^2$. The main tool to study spectral properties of PDEs with periodic coefficients is Floquet-Bloch theory [105, 108, 45, 85]. A Bloch solution to (2.3) is a non-zero function of the form

$$E_3(\vec{x}) = \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}} u(\vec{x}), \tag{2.4}$$

where u is a periodic function with respect to $\mathcal{L}$ and $\vec{k} \in \Omega^* = (-\pi,\pi]^2$ denotes the wave vector. By inserting the Bloch ansatz (2.4), the original problem (2.3) posed on the infinite domain $\mathbb{R}^2$ becomes

$$-(\nabla + \mathrm{i}\vec{k}) \cdot (\nabla + \mathrm{i}\vec{k}) u(\vec{x}) = \left(\frac{\omega}{c}\right)^2 \epsilon(\vec{x},\omega) u(\vec{x}), \tag{2.5}$$

which is a family (parameterized by the wave vector $\vec{k}$) of eigenvalue problems in ω on the unit cell Ω with periodic boundary conditions. When calculating the dispersion relations $\omega(\vec{k})$ numerically, frequently only a selection of wave vectors $\vec{k}$ along the line segments between the points Γ, X, and M, as shown in Figure 2.1, is considered. The triangular path formed by these points is called the boundary of the irreducible Brillouin zone [70].

We assume that electromagnetic energy may be transferred into the material, but energy is not transferred from the material into the electromagnetic field. In other words, the material is assumed to be passive [28, 64], which corresponds to the condition

$$\omega\epsilon(\omega) \in \mathbb{C}_+ := \{z \in \mathbb{C} : 0 \leq \arg z < \pi,\ z \neq 0\} \quad \forall \omega \in \mathbb{C}_+ \tag{2.6}$$

An important consequence of condition (2.6) is that $\operatorname{Im}\epsilon(\omega) \geq 0$ for $\omega > 0$. The spectral problem (2.5) with a passive material model was analyzed in [42], where it was proved that the spectrum is discrete.

In principle, any general space-dependent permittivity can be handled when the eigenvalue problem (2.5) is discretized with finite elements. However, for simplicity, we restrict the permittivity to be piecewise constant with respect to a finite partitioning $\Omega = \Omega_1 \cup \cdots \cup \Omega_S$ of the unit cell,

$$\epsilon(\vec{x}, \omega) = \sum_{s=1}^{S} \epsilon_s(\omega) \chi_{\Omega_s}(\vec{x}). \tag{2.7}$$

Here, χ_{Ω_s} designates the indicator function for the subdomain Ω_s. A finite-element discretization then yields a matrix eigenvalue problem of the form

$$K(\vec{k})\vec{u} - \omega^2 \sum_{s=1}^{S} \epsilon_s(\omega) M_s \vec{u} = 0. \tag{2.8}$$

If the permittivities $\epsilon_1, \ldots, \epsilon_S$ of all constituents are independent of the frequency ω, the eigenvalue problem is linear in ω^2. In the frequency-dependent case, however, the permittivities $\epsilon_1, \ldots, \epsilon_S$ are, in general, complicated unknown functions in the frequency ω. It is therefore common to model this frequency dependency by fitting a rational function satisfying the passivity condition (2.6) to measurement data. The resulting Lorentz model takes the form

$$\epsilon_s(\omega) = \alpha_s + \sum_{j=1}^{L_s} \frac{\xi_{s,j}^2}{\eta_{s,j}^2 - \omega^2 - \mathrm{i}\gamma_{s,j}\omega}$$

and leads to a rational eigenvalue problem when inserted into (2.8).

2.3 Boundary-element discretizations

As a model problem for this section, we will consider the Laplace eigenvalue problem with homogeneous Dirichlet boundary conditions,

$$\begin{aligned} -\Delta u &= \lambda u \qquad && \text{in } \Omega \subset \mathbb{R}^3, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{2.9}$$

on some domain $\Omega \subset \mathbb{R}^3$ with a locally Lipschitz continuous boundary $\partial\Omega$. A classical finite-element discretization of the problem (2.9) leads to a generalized linear eigenvalue problem of the form

$$Kx = \lambda Mx,$$

which can be solved by established techniques. Here, K and M denote the stiffness and mass matrices of the problem, respectively, and x represents the discretized eigenfunction u. Unfortunately, finite-element methods require a discretization of the domain Ω, which may lead to inconveniences, especially if the domain is unbounded.

With boundary-element discretizations, on the other hand, only a discretization of the boundary $\partial\Omega$ is needed. In [124], a boundary integral formulation for the model problem (2.9) is derived in the following way. Rewriting the eigenvalue problem as a homogeneous Helmholtz equation, the Helmholtz representation formula implies

$$u(t) = \frac{1}{4\pi} \int_{\partial\Omega} \frac{\mathrm{e}^{\mathrm{i}\sqrt{\lambda}\|t-\eta\|}}{\|t-\eta\|} u_n(\eta) \, \mathrm{d}S(\eta), \tag{2.10}$$

where $u_n \in H^{-\frac{1}{2}}(\partial\Omega)$ is the exterior normal derivative of u. A Galerkin approach to weakly enforce the zero Dirichlet boundary conditions on u therefore leads to

$$\frac{1}{4\pi}\int_{\partial\Omega} v(\xi)\int_{\partial\Omega}\frac{\mathrm{e}^{\mathrm{i}\sqrt{\lambda}\|\xi-\eta\|}}{\|\xi-\eta\|}u_n(\eta)\,\mathrm{d}S(\eta)\,\mathrm{d}S(\xi) = 0$$

with a test function $v \in H^{-\frac{1}{2}}(\partial\Omega)$. Discretizing u_n and v by piecewise constant functions with respect to some triangulation $\{\triangle_1, \dots, \triangle_n\}$ of the boundary $\partial\Omega$ results in a nonlinear eigenvalue problem of the form (1.1), where the (i,j)-th entry of the matrix-valued function T is given by

$$[T(\lambda)]_{ij} = \frac{1}{4\pi}\int_{\triangle_i}\int_{\triangle_j}\frac{\mathrm{e}^{\mathrm{i}\sqrt{\lambda}\|\xi-\eta\|}}{\|\xi-\eta\|}\,\mathrm{d}S(\eta)\,\mathrm{d}S(\xi). \tag{2.11}$$

The eigenvectors of this problem represent the discretized Neumann data u_n of an (approximate) eigenfunction for the Laplace eigenvalue problem (2.9), which can be reconstructed by means of the representation formula (2.10).

Since the eigensystem of the model problem is known to be real, the imaginary part of the matrix-valued function T is frequently omitted, resulting in a nonlinear eigenvalue problem with entries

$$[\operatorname{Re} T(\lambda)]_{ij} = \frac{1}{4\pi}\int_{\triangle_i}\int_{\triangle_j}\frac{\cos(\sqrt{\lambda}\|\xi-\eta\|)}{\|\xi-\eta\|}\,\mathrm{d}S(\eta)\,\mathrm{d}S(\xi). \tag{2.12}$$

The formulation (2.12) is also obtained when applying the Multiple Reciprocity Method [104, 72], which is commonly used in practice, to the model problem (2.9). However, the omission of the imaginary part has a tendency to introduce spurious eigenvalues; see, e.g., [30].

Evaluating the matrix entries in (2.11) or (2.12) is rather expensive due to the non-locality and singularity of the integral kernel; compare [115, 123]. Some work may be saved by additively splitting the kernel into a singular and a nonsingular component. The singular component may be chosen independent of λ so that its integral can be precomputed, leaving only a nonsingular integration to be done for each individual λ.

To eliminate the need for repeated integration, many authors [72, 78, 79, 104] consider a truncated Taylor expansion of the integrand, leading to a polynomial approximation

$$T(\lambda) \approx T_0 + \lambda T_1 + \lambda^2 T_2 + \cdots \tag{2.13}$$

of the matrix-valued function T. For other, potentially more effective polynomial approximations, see [130] as well as Chapter 5.

As a more realistic application, we consider time-harmonic vibrations of a three-dimensional, elastic shell-like structure Ω_S, surrounded by a compressible fluid Ω_F, which are governed by the following eigenvalue problem (cf. [62, Section 1.3]): Find $(\mathbf{u}, p, \omega) \in \mathbf{H}^1(\Omega_s) \times H^1_{\text{loc}}(\Omega_F) \times \mathbb{C}$ such that $(\mathbf{u}, p) \neq (0,0)$ and

$$\begin{aligned}
-\varrho_S\omega^2\mathbf{u} - \mu\Delta u - (\lambda+\mu)\operatorname{grad}\operatorname{div}\mathbf{u} &= 0 && \text{in } \Omega_S,\\
-\Delta p - \left(\frac{\omega}{c_F}\right)^2 p &= 0 && \text{in } \Omega_F,\\
\mathcal{T}\mathbf{u} &= 0 && \text{on } \Gamma_1,\\
\rho_F\omega^2\mathbf{n}\cdot\mathbf{u} &= \frac{\partial p}{\partial \mathbf{n}} && \text{on } \Gamma_0,\\
-p\mathbf{n} &= \mathcal{T}\mathbf{u} && \text{on } \Gamma_0.
\end{aligned}$$

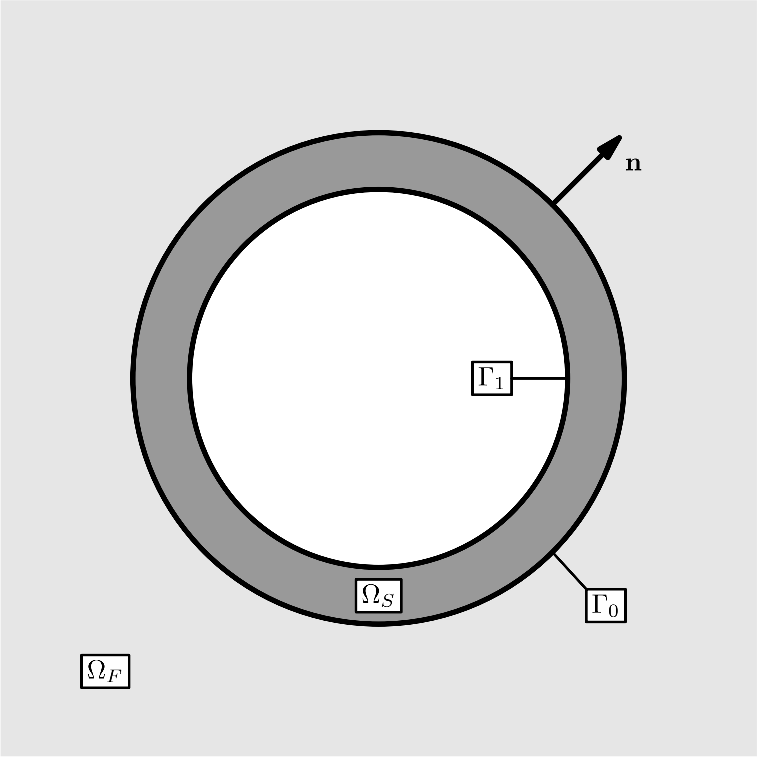

Figure 2.2: A three-dimensional, elastic shell-like structure Ω_S surrounded by a compressible fluid Ω_F.

Here, Γ_0 and Γ_1 are the outer and inner boundary of the structure Ω_S, respectively, and $\mathbf{n}$ is the unit normal vector on Γ_0 pointing into Ω_F; see Figure 2.2 for an illustration. Furthermore, ρ_S and ρ_F are the densities of the structure and the fluid, c_F is the speed of sound in the fluid, λ and μ are the Lamé parameters describing the elasticity of the structure, and $\mathcal{T}$ denotes the boundary stress operator. The eigenvalue ω represents the angular frequency of the vibration. The eigenvector consists of the displacement field $\mathbf{u}$ for the shell as well as the acoustic pressure p in the fluid. Since the fluid occupies the unbounded exterior region of Ω_S, we additionally impose an outgoing radiation condition in the sense of [114, Chapter 8, Definition 1.5] on p to make the problem well-posed.

Discretizing the differential equation for the shell Ω_S by finite elements and the differential equation for the fluid Ω_F by the boundary-element approach described above yields a nonlinear eigenvalue problem of the form (1.1); see also [41, 110]. The interesting eigenvalues are those with positive real part on or close to the real axis.

2.4 Vibrating string with elastically attached mass

The analysis of eigenvibrations for mechanical structures with elastically attached masses frequently leads to nonlinear eigenvalue problems of the form (1.1); see,

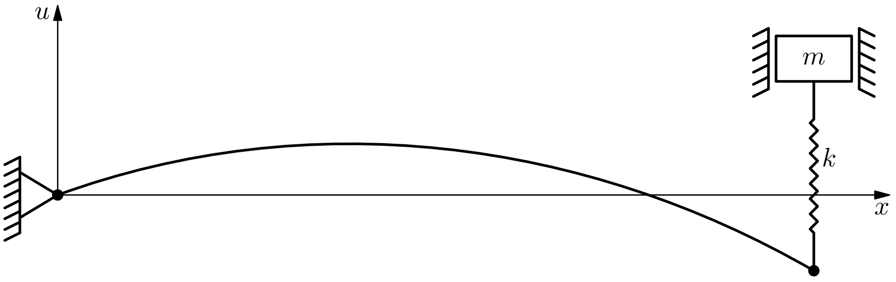

Figure 2.3: Illustration of a vibrating string, which is clamped at the left end and has a vertically moving mass attached to the right end via an elastic spring.

e.g., [144, 135, 121, 143]. As a model problem, we consider here a limp string of unit length, which is clamped at one end. The other end is free but has a mass m attached to it via an elastic spring of stiffness k. The movement of the mass is constrained to the vertical direction, whereas the string, in its equilibrium state, is horizontal; see Figure 2.3 for an illustration. Under these circumstances, the eigenvibrations of the string are governed by the eigenvalue problem [122]

$$-u''(x) = \lambda u(x), \quad u(0) = 0, \quad u'(1) + \phi(\lambda)u(1) = 0, \quad \phi(\lambda) = \frac{\lambda\eta m}{\lambda - \eta}, \tag{2.14}$$

where u denotes the displacement and $\eta = \frac{k}{m}$. Furthermore, it is assumed that the clamped end resides at $x = 0$ and the free end with the mass at $x = 1$.

The eigenproblem (2.14) is simple enough to admit a semi-analytic solution. One easily calculates that the differential equation together with the boundary condition at $x = 0$ implies

$$u(x) = C \cdot \sin\bigl(\sqrt{\lambda} \cdot x\bigr), \quad C \in \mathbb{R}. \tag{2.15}$$

Inserting this expression into the boundary condition at $x = 1$ and rearranging shows that λ is an eigenvalue of (2.14) if and only if

$$\tan\bigl(\sqrt{\lambda}\bigr) = \frac{1}{m\sqrt{\lambda}} - \frac{\sqrt{\lambda}}{k}.$$

Solving the above equation numerically, we thus obtain all eigenfrequencies λ of the string. The corresponding eigenmodes are then given by (2.15).

Discretizing the eigenvalue problem (2.14) with finite elements on the uniform grid $\bigl\{x_i = \frac{i}{n} : i = 1, \dots, n\bigr\}$ of size $h = \frac{1}{n}$ yields a nonlinear matrix eigenvalue problem of the form

$$\bigl(K(\lambda) - \lambda M\bigr)v = 0 \tag{2.16}$$

with $K(\lambda) = K_0 + \phi(\lambda)e_n e_n^{\mathsf{T}}$ and

$$K_0 = \frac{1}{h}\begin{bmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & 2 & -1 \\ & & -1 & 1 \end{bmatrix}, \quad M = \frac{h}{6}\begin{bmatrix} 4 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & 4 & 1 \\ & & 1 & 2 \end{bmatrix}.$$

The matrices K_0 and M are positive definite and $e_n e_n^\mathsf{T}$ is positive semidefinite. Consequently, we have $\kappa := v^\mathsf{T} K_0 v > 0$, $\mu := v^\mathsf{T} M v > 0$, and $\epsilon := v^\mathsf{T}(e_n e_n^\mathsf{T})v \geq 0$ for any non-zero vector $v \in \mathbb{R}^n$. Furthermore, one readily verifies that the function

$$f_v : (\eta, \infty) \to \mathbb{R}, \quad f_v(\lambda) = v^\mathsf{T}\big(K(\lambda) - \lambda M\big)v = \kappa + \epsilon\phi(\lambda) - \mu\lambda$$

is strictly decreasing for $\lambda \in (\eta, \infty)$, has positive values in the vicinity of η, and satisfies $\lim\limits_{\lambda \to \infty} f_v(\lambda) = -\infty$. Therefore, the equation $f_v(\lambda) = 0$ has exactly one solution for all non-zero vectors $v \in \mathbb{R}^n$, showing that the nonlinear eigenvalue problem (2.16) admits a Rayleigh functional ρ with $D(\rho) = \mathbb{R}^n \setminus \{0\}$ on the interval (η, ∞); compare Chapter 7. Hence, there exists a variational characterization for the eigenvalues $\lambda > \eta$ of problem (2.16).

Chapter 3

Minimal invariant pairs

Minimal invariant pairs have been introduced in [83] and generalize the notion of invariant subspaces to the nonlinear setting. They provide a means to represent a portion of the eigensystem of a nonlinear eigenvalue problem in a numerically robust way and hence constitute a vital component of the methods presented in the upcoming chapters as well as many other methods aiming at the computation of several eigenpairs. In this chapter, we will first give a short exposition of the existing theory and later introduce some new developments.

3.1 Basic principles

Originally, invariant pairs have been defined in [83] for the special case that the matrix valued function T defining the nonlinear eigenvalue problem is given as a sum of constant coefficient matrices $T_1, \dots, T_d \in \mathbb{C}^{n \times n}$, weighted by scalar, holomorphic functions $f_1, \dots, f_d$,

$$T(\lambda) = f_1(\lambda)T_1 + \cdots + f_d(\lambda)T_d. \tag{3.1}$$

A pair $(X, \Lambda) \in \mathbb{C}^{n \times m} \times \mathbb{C}^{m \times m}$ is then called invariant if $\mathbf{T}(X, \Lambda) = 0$, where

$$\mathbf{T}(X, \Lambda) = T_1 X f_1(\Lambda) + \cdots + T_d X f_d(\Lambda). \tag{3.2}$$

Here, $f_j(\Lambda)$ is to be understood as a matrix function in the sense of [60].

In this work, we will favor a different characterization of invariant pairs as it leads to conceptually more elegant proofs and does not require a specific form of the matrix-valued function, such as (3.1). The alternative characterization can be derived as follows. Since the matrix-valued function T is assumed to be holomorphic, we rewrite the nonlinear eigenvalue problem (1.1) using the Cauchy integral formula as

$$T(\lambda)x = \frac{1}{2\pi \mathrm{i}} \int_{\partial B_\delta(\lambda)} T(\xi)x(\xi - \lambda)^{-1} \, \mathrm{d}\xi = 0, \tag{3.3}$$

where $B_\delta(\lambda)$ denotes the closed ball of radius δ around λ and $\delta > 0$ is chosen such that $B_\delta(\lambda) \subset \mathcal{D}$. The formulation (3.3) of the nonlinear eigenvalue problem has the advantage that it naturally extends into a block version. Hence, we are led to the subsequent definition.

Definition 3.1.1. For any pair $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$, we call

$$\mathbf{T}(X, \Lambda) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma T(\xi) X (\xi I - \Lambda)^{-1} \, \mathrm{d}\xi \tag{3.4}$$

the *block residual* associated with (X, Λ). Moreover, (X, Λ) is called an *invariant pair* of the nonlinear eigenvalue problem (1.1) if and only if the associated block residual vanishes. Here, Γ is a contour, i.e., a simply closed curve in $\mathcal{D}$, enclosing the eigenvalues of Λ in its interior. Note that because the eigenvalues of Λ are isolated points in $\mathcal{D}$, such a contour Γ always exists.

Proposition 3.1.2. *Assume that the matrix-valued function of the nonlinear eigenvalue problem* (1.1) *is of the form* (3.1). *Then for any pair* (X, Λ)*, the block residual satisfies* (3.2).

Proof. Let Γ be a contour enclosing the eigenvalues of Λ in its interior. By the contour integral representation of matrix functions [60, Definition 1.11], we have

$$f_j(\Lambda) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma f_j(\xi) (\xi I - \Lambda)^{-1} \, \mathrm{d}\xi, \quad j = 1, \dots, d.$$

Hence, inserting the representation (3.1) of the matrix-valued function T into the contour integral definition (3.4) of the block residual gives

$$\begin{aligned} \mathbf{T}(X, \Lambda) &= \frac{1}{2\pi \mathrm{i}} \int_\Gamma T(\xi) X (\xi I - \Lambda)^{-1} \, \mathrm{d}\xi = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \sum_{j=1}^{d} f_j(\xi) T_j X (\xi I - \Lambda)^{-1} \, \mathrm{d}\xi \\ &= \sum_{j=1}^{d} T_j X \frac{1}{2\pi \mathrm{i}} \int_\Gamma f_j(\xi) (\xi I - \Lambda)^{-1} \, \mathrm{d}\xi = T_1 X f_1(\Lambda) + \cdots + T_d X f_d(\Lambda) \end{aligned}$$

as claimed. □

Proposition 3.1.2 shows the original characterization of invariance is equivalent to ours whenever the matrix-valued function T is of the special form (3.1). In particular, Equations (3.2) and (3.4) define the same block residual. This is very convenient because despite the theoretical elegance of the latter, the former is more amenable to actual computation. Therefore, we will resort to (3.2) for calculating block residuals in most situations. In the event that the use of (3.2) fails, one can avoid contour integration in the evaluation of block residuals by employing a Parlett-Schur-type algorithm [32] incorporating block diagonalization and Taylor expansion.

As is easily seen, an invariant pair (X, Λ) with $m = 1$ (i.e., X is a vector and Λ a scalar) amounts to an eigenpair, provided that $X \neq 0$. In case $m > 1$, a similar connection can be established [83, Lemma 4 (2.)] in that for every eigenpair (u, λ) of Λ, (Xu, λ) is an eigenpair of the nonlinear eigenvalue problem (1.1), provided that $Xu \neq 0$. The latter condition can be rephrased as requiring that $u = 0$ be the only null vector of the augmented matrix

$$\begin{bmatrix} X \\ \Lambda - \lambda I \end{bmatrix}. \tag{3.5}$$

This justifies the following definition.

Definition 3.1.3. A pair $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ is called *minimal* if and only if the matrix in (3.5) has full column rank for all eigenvalues λ of Λ or, equivalently, for all $\lambda \in \mathbb{C}$.

Notice that the original definition of minimality in [83, Definition 2] differs from the one given above. However, both definitions are equivalent as borne out by the subsequent proposition.

Proposition 3.1.4. *The pair $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ is minimal if and only if there exists an integer ℓ such that the matrix*

$$\mathbf{V}_\ell(X,\Lambda) := \begin{bmatrix} X \\ X\Lambda \\ \vdots \\ X\Lambda^{\ell-1} \end{bmatrix} \tag{3.6}$$

has full column rank.

Proof. The statement of the lemma is a well-known result in mathematical systems theory (see, e.g., [71, Theorem 2.4-9 (2.)]), where the stated equivalence is exploited in the Hautus test [57] for observability of a matrix pair. In this setting, (3.5) and (3.6) correspond to the Hautus and Kalman observability matrices, respectively. □

Definition 3.1.5 ([83, Definition 2])**.** Let (X, Λ) be a minimal pair. The smallest integer ℓ for which the matrix $\mathbf{V}_\ell(X, \Lambda)$ in (3.6) has full column rank is called the *minimality index* of (X, Λ) and we say that (X, Λ) is minimal of index ℓ.

In this work, we will make regular use of both characterizations of minimality, depending on which is more utile for the situation at hand. The criterion given in Proposition 3.1.4 has the slight disadvantage that it requires knowledge of (an upper bound on) the minimality index of the pair (X, Λ) under consideration.

3.1.1 Bounds on the minimality index. To bound the minimality index of a pair $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$, we exploit that the row space of the matrix $\mathbf{V}_\ell(X, \Lambda)$ in (3.6), i.e., the range of $\mathbf{V}_\ell(X, \Lambda)^\mathsf{T}$, is the block Krylov subspace $\mathcal{K}_\ell(\Lambda^\mathsf{T}, X^\mathsf{T})$. The well-known termination property of Krylov subspaces then gives rise to the following result.

Proposition 3.1.6. *Let $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ and denote, for $j = 1, 2, \ldots$, the rank of the matrix $\mathbf{V}_j(X, \Lambda)$ by r_j. Let ℓ be the smallest integer such that $r_\ell = r_{\ell+1}$. Then,*

$$r_1 < \cdots < r_\ell = r_{\ell+1} = r_{\ell+2} = \cdots \,. \tag{3.7}$$

Furthermore, if (X, Λ) is minimal, then its minimality index is equal to ℓ.

Proof. First of all, we will demonstrate that ℓ is well-defined. By definition, the matrices $\mathbf{V}_j(X, \Lambda)$ satisfy the recursion $\mathbf{V}_{j+1}(X, \Lambda) = \begin{bmatrix} \mathbf{V}_j(X,\Lambda) \\ X\Lambda^j \end{bmatrix}$, showing that their row spaces are nested:

$$\operatorname{span} \mathbf{V}_j(X, \Lambda)^\mathsf{T} \subseteq \operatorname{span} \mathbf{V}_{j+1}(X, \Lambda)^\mathsf{T}, \quad j = 1, 2, \ldots \,.$$

Therefore, and since r_j equals the dimension of $\operatorname{span} \mathbf{V}_j(X, \Lambda)^\mathsf{T}$, the sequence (r_j) is non-decreasing. On the other hand, since $\mathbf{V}_j(X, \Lambda)$ has m columns, r_j cannot exceed m. Consequently, there are indices j for which $r_j = r_{j+1}$, and because $j \geq 1$, there is also a smallest one.

The inequality part of (3.7) then follows from the definition of ℓ. The equality part of (3.7) is proved by induction. To this end, assume that $r_\ell = r_{\ell+1} = \cdots = r_j$, which is true for $j = \ell + 1$ by the definition of ℓ. This assumption implies that the nested row spaces $\operatorname{span} \mathbf{V}_\ell(X, \Lambda)^\mathsf{T}, \ldots, \operatorname{span} \mathbf{V}_j(X, \Lambda)^\mathsf{T}$ are, in fact, identical. Hence, again using the definition of $\mathbf{V}_j(X, \Lambda)$,

$$\begin{aligned} \operatorname{span} \mathbf{V}_{j+1}(X, \Lambda)^\mathsf{T} &= \operatorname{span}\left[\begin{smallmatrix} X \\ \mathbf{V}_j(X,\Lambda)\cdot\Lambda \end{smallmatrix}\right]^\mathsf{T} = \operatorname{span} X^\mathsf{T} + \operatorname{span}\bigl(\Lambda^\mathsf{T} \mathbf{V}_j(X, \Lambda)^\mathsf{T}\bigr) \\ &= \operatorname{span} X^\mathsf{T} + \Lambda^\mathsf{T} \operatorname{span} \mathbf{V}_j(X, \Lambda)^\mathsf{T} = \operatorname{span} X^\mathsf{T} + \Lambda^\mathsf{T} \operatorname{span} \mathbf{V}_\ell(X, \Lambda)^\mathsf{T} \\ &= \operatorname{span}\left[\begin{smallmatrix} X \\ \mathbf{V}_\ell(X,\Lambda)\cdot\Lambda \end{smallmatrix}\right]^\mathsf{T} = \operatorname{span} \mathbf{V}_{\ell+1}(X, \Lambda)^\mathsf{T} = \operatorname{span} \mathbf{V}_\ell(X, \Lambda)^\mathsf{T}, \end{aligned}$$

showing that $r_{j+1} = r_\ell$, which finishes the induction.

The statement about the minimality index is a direct consequence of (3.7) if we can show that minimality of (X, Λ) entails $r_\ell = m$. The latter is true because $r_\ell < m$ implies $r_j < m$ for all $j = 1, 2, \ldots$ by virtue of (3.7), contradicting the minimality of (X, Λ). □

Proposition 3.1.6 readily leads to an upper bound on the minimality index.

Corollary 3.1.7. *Let the pair* $(X, \Lambda) \in \mathbb{C}^{n \times m} \times \mathbb{C}^{m \times m}$ *be minimal of index* ℓ*. Then,* $\ell \leq m$*.*

Proof. Let r_j, $j = 1, 2, \ldots$ be defined as in Proposition 3.1.6. By a simple induction, Inequality (3.7) yields $r_1 \leq r_\ell - \ell + 1$. Moreover, as shown in the proof of Proposition 3.1.6, the minimality of the pair (X, Λ) entails $r_\ell = m$. Assuming $\ell > m$, we therefore obtain $r_1 < 1$, or $r_1 = 0$ since r_1 is a non-negative integer. The proof is now concluded by noticing that $r_1 = 0$ implies $0 = \mathbf{V}_1(X, \Lambda) = X$ and, thus, $\mathbf{V}_j(X, \Lambda) = 0$ for all $j = 1, 2, \ldots$, contradicting the minimality of (X, Λ). □

Corollary 3.1.7 has originally been proved in [83, Lemma 5] by means of the Cayley-Hamilton theorem; the proof given here is due to the author. A stronger upper bound on the minimality index has been given in [19, Lemma 5.1]. There, the minimality index of a pair (X, Λ) is shown to be smaller than the cumulated size of the largest Jordan blocks belonging to the distinct eigenvalues of Λ. Lower bounds on the minimality index can also be derived. A simple example taken from [19] is given below.

Proposition 3.1.8. *Let the pair* $(X, \Lambda) \in \mathbb{C}^{n \times m} \times \mathbb{C}^{m \times m}$ *be minimal of index* ℓ*. Then,* $\ell \geq \lceil \frac{m}{n} \rceil$*, where* $\lceil \cdot \rceil$ *indicates rounding to the next larger integer.*

Proof. Since the matrix $\mathbf{V}_\ell(X, \Lambda)$ is of size $\ell n \times m$, it cannot have full column rank unless $\ell n \geq m$. Solving for ℓ and taking into account that ℓ needs to be an integer verifies the assertion. □

3.1.2 Connection to the eigensystem.

Minimal invariant pairs (X, Λ) whose Λ-component is in Jordan canonical form are closely related to the concept of a Jordan pair studied in [47]. Jordan pairs, in turn, are intimately connected to the eigensystem of the underlying nonlinear eigenvalue problem. To describe this relationship, we need to introduce the notion of a Jordan chain, which is most elegantly defined in terms of root functions; see also [98].

Definition 3.1.9. A holomorphic, vector-valued function $x : \mathcal{D} \to \mathbb{C}^n$ is called a *root function* of the matrix-valued function T at $\lambda_0 \in \mathcal{D}$ if and only if

$$T(\lambda_0)x(\lambda_0) = 0, \quad x(\lambda_0) \neq 0.$$

The order of the zero of $\lambda \mapsto T(\lambda)x(\lambda)$ at $\lambda = \lambda_0$ is called the *multiplicity* of the root function x.

Definition 3.1.10. A sequence of m vectors, $x_0, x_1, \dots, x_{m-1}$, is called a *Jordan chain* of the matrix-valued function T at λ_0, consisting of the *eigenvector* x_0 and the *generalized eigenvectors* $x_1, \dots, x_{m-1}$, if and only if the function defined by

$$x(\lambda) = \sum_{j=0}^{m-1} (\lambda - \lambda_0)^j x_j$$

constitutes a root function of T at λ_0 of multiplicity at least m. The integer m is called the *length* of the Jordan chain.

It is important to note that for linear eigenvalue problems, $T(\lambda) = \lambda I - A$, Definition 3.1.10 coincides with the usual notions of (generalized) eigenvectors and Jordan chains. Using these definitions, we can now formulate the following equivalence.

Lemma 3.1.11 ([47, Lemma 2.1]). *Let λ_0 be an eigenvalue of the matrix-valued function T and consider a matrix $X = [x_0, \dots, x_{m-1}] \in \mathbb{C}^{n \times m}$ with $x_0 \neq 0$. Then $x_0, \dots, x_{m-1}$ is a Jordan chain of T at λ_0 if and only if $(X, J_m(\lambda_0))$ constitutes an invariant pair of T, where $J_m(\lambda_0)$ denotes a Jordan block of order m associated with λ_0.*

There may be multiple Jordan chains belonging to the same eigenvalue. In fact, for every eigenvalue λ_0 of a matrix-valued function T, there exists a *canonical system of (generalized) eigenvectors*. Via Definition 3.1.10, this system is linked to a *canonical system of root functions* $x^{(1)}(\cdot), \dots, x^{(g)}(\cdot)$ whose multiplicities sum up to the algebraic multiplicity of λ_0. Furthermore, their number, g, equals the geometric multiplicity of λ_0, and $x^{(1)}(\lambda_0), \dots, x^{(g)}(\lambda_0)$ are linearly independent eigenvectors. The statement of Lemma 3.1.11 can be generalized to the case of multiple Jordan chains.

Proposition 3.1.12. *Let λ_0 be an eigenvalue of the matrix-valued function T and consider a matrix*

$$X = \bigl[X^{(1)}, \dots, X^{(g)}\bigr], \quad X^{(j)} = \bigl[x_0^{(j)}, \dots, x_{m_j-1}^{(j)}\bigr] \in \mathbb{C}^{n \times m_j}$$

with $x_0^{(j)} \neq 0$ for all $j = 1, \dots, g$. Then $x_0^{(j)}, \dots, x_{m_j-1}^{(j)}$ is a Jordan chain of T at λ_0 for every $j = 1, \dots, g$ if and only if (X, J_{λ_0}) constitutes an invariant pair of T, where $J_{\lambda_0} = \operatorname{diag}\{J_{m_1}(\lambda_0), \dots, J_{m_g}(\lambda_0)\}$ and $J_{m_1}(\lambda_0), \dots, J_{m_g}(\lambda_0)$ denote Jordan blocks of orders $m_1, \dots, m_g$ associated with λ_0. Furthermore, (X, J_{λ_0}) is minimal if and only if the vectors $x_0^{(1)}, \dots, x_0^{(g)}$ are linearly independent.

Proof. By Proposition 3.3.1 below, (X, J_{λ_0}) is invariant if and only if $\bigl(X^{(j)}, J_{m_j}(\lambda_0)\bigr)$ is invariant for all $j = 1, \dots, g$. The first statement therefore follows by applying Lemma 3.1.11 to each of the pairs $\bigl(X^{(j)}, J_{m_j}(\lambda_0)\bigr)$.

For the second statement, we will employ the characterization of minimality in Definition 3.1.3. So assume that $x_0^{(1)}, \dots, x_0^{(g)}$ are linearly independent and let $u = [u_1^\mathsf{T}, \dots, u_g^\mathsf{T}]^\mathsf{T}$, partitioned in accordance with the block structure of X and J_{λ_0}, be an arbitrary vector satisfying

$$\begin{bmatrix} X \\ J_{\lambda_0} - \lambda_0 I \end{bmatrix} u = 0. \tag{3.8}$$

The second block row of the above equation shows that $J_{m_j}(\lambda_0) u_j = \lambda_0 u_j$ for $j = 1, \dots, g$. Since $J_{m_j}(\lambda_0)$ is a Jordan block, the latter implies $u_j = \alpha_j e_1$ for some suitable $\alpha_j \in \mathbb{C}$, where e_1 denotes the first unit vector of appropriate size. Inserting the last equation into the first block row of Equation (3.8) yields

$$\alpha_1 x_0^{(1)} + \cdots \alpha_g x_0^{(g)} = 0, \tag{3.9}$$

implying $\alpha_1 = \cdots = \alpha_g = 0$ due to the linear independence of $x_0^{(1)}, \dots, x_0^{(g)}$. Thus, $u = 0$ is the only solution of Equation (3.8), which establishes the minimality of (X, J_{λ_0}).

Conversely, assume (X, J_{λ_0}) to be minimal and let $\alpha_1, \dots, \alpha_g \in \mathbb{C}$ satisfy Equation (3.9). Then, $u = [\alpha_1 e_1^\mathsf{T}, \dots, \alpha_g e_1^\mathsf{T}]^\mathsf{T}$ solves Equation (3.8), which entails $u = 0$ by minimality of (X, J_{λ_0}). Thus, $\alpha_1 = \cdots = \alpha_g = 0$, confirming the linear independence of $x_0^{(1)}, \dots, x_0^{(g)}$. □

Proposition 3.1.12 represents a slight extension of [47, Theorem 2.3], where it is assumed that the Jordan chains contained in the matrix X constitute a full canonical system of (generalized) eigenvectors. The result may be further extended by removing the restriction to one eigenvalue only.

Theorem 3.1.13. *Let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of the matrix-valued function T and consider a matrix*

$$X = \bigl[X^{(1)}, \dots, X^{(k)}\bigr], \quad X^{(i)} = \bigl[X^{(i,1)}, \dots, X^{(i,g_i)}\bigr],$$
$$X^{(i,j)} = \bigl[x_0^{(i,j)}, \dots, x_{m_{i,j}-1}^{(i,j)}\bigr] \in \mathbb{C}^{n \times m_{i,j}}$$

with $x_0^{(i,j)} \neq 0$ for all $i = 1, \dots, k$ and $j = 1, \dots, g_i$. Then $x_0^{(i,j)}, \dots, x_{m_{i,j}-1}^{(i,j)}$ is a Jordan chain of T at λ_i for every $i = 1, \dots, k$ and $j = 1, \dots, g_i$ if and only if (X, J) constitutes an invariant pair of T, where

$$J = \operatorname{diag}\{J_{\lambda_1}, \dots, J_{\lambda_k}\}, \quad J_{\lambda_i} = \operatorname{diag}\{J_{m_{i,1}}(\lambda_i), \dots, J_{m_{i,g_i}}(\lambda_i)\},$$

and $J_{m_{i,j}}(\lambda_i)$ denotes a Jordan block of order $m_{i,j}$ associated with λ_i for $i = 1, \dots, k$ and $j = 1, \dots, g_i$. Furthermore, (X, J) is minimal if and only if for each $i = 1, \dots, k$, the vectors $x_0^{(i,1)}, \dots, x_0^{(i,g_i)}$ are linearly independent.

Proof. By Propositions 3.3.1 and 3.3.2 further down, the pair (X, J) is minimal invariant if and only if each of the pairs $(X^{(i)}, J_{\lambda_i})$, $i = 1, \dots, k$ is minimal invariant. The statement of the theorem therefore follows by applying Proposition 3.1.12 to each of these pairs. □

Theorem 3.1.13 secures that an arbitrary portion of the eigensystem of a matrix-valued function T can be represented as a minimal invariant pair. For the opposite direction of extracting spectral information from a minimal invariant pair (X, Λ), one exploits that $(XG, G^{-1}\Lambda G)$ is again a minimal invariant pair for any invertible matrix G of appropriate size by Proposition 3.2.3 below. If G is chosen such that $G^{-1}\Lambda G$ is in Jordan canonical form, Theorem 3.1.13 implies that XG contains Jordan chains. Thus, every minimal invariant pair replicates a part of the eigensystem of the corresponding matrix-valued function. Unfortunately, though, Jordan chains are fragile under perturbations and, hence, not well suited for numerical purposes; see [141] for a recent discussion. In a computational setting, it is therefore advisable to replace the Jordan canonical form by a numerically more stable transformation, such as a reduction to Schur form.

The (lack of) completeness of the spectral information supplied by a minimal invariant pair is measured by a positive integer, called the multiplicity of the pair. A pair (X, Λ) has multiplicity one if it holds a full canonical system of (generalized) eigenvectors for every eigenvalue of Λ. Each missing (generalized) eigenvector increments the multiplicity, resulting in the subsequent definition.

Definition 3.1.14. To any minimal invariant pair (X, Λ), we assign the *multiplicity*

$$1 + \sum_{\lambda \in \operatorname{spec} \Lambda} (\operatorname{alg}_T \lambda - \operatorname{alg}_\Lambda \lambda).$$

(X, Λ) is called *simple* if its multiplicity is 1 and *multiple* if it has higher multiplicity.

3.2 Characterizing minimality using non-monomial bases

Typically, the minimality index of a minimal pair $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ is quite small. Generically, it is equal to one unless $m > n$; compare Proposition 3.1.8. For larger minimality indices, though, the monomials Λ^i, $i = 0, \dots, \ell - 1$ within the matrix $\mathbf{V}_\ell(X, \Lambda)$ may cause numerical instabilities. To mitigate this effect, we propose replacing the monomials by a different polynomial basis. That is, instead of $\mathbf{V}_\ell(X, \Lambda)$, we consider the matrix

$$\mathbf{V}_\ell^p(X, \Lambda) = \begin{bmatrix} Xp_0(\Lambda) \\ \vdots \\ Xp_{\ell-1}(\Lambda) \end{bmatrix}, \tag{3.10}$$

where the superscript p indicates the use of a basis $p_0, \dots, p_{\ell-1}$ for the vector space of polynomials of degree at most $\ell - 1$. Note that $\mathbf{V}_\ell^p(X, \Lambda)$ includes $\mathbf{V}_\ell(X, \Lambda)$ as a special case if $p_0, \dots, p_{\ell-1}$ is the monomial basis, $p_i(\lambda) = \lambda^i$, $i = 0, \dots, \ell - 1$. Intuitively, switching the polynomial basis amounts to an invertible recombination of the blocks in the matrix (3.10).

Proposition 3.2.1. *Let $p_0, \dots, p_{\ell-1}$ and $\tilde{p}_0, \dots, \tilde{p}_{\ell-1}$ be two bases for the vector space of polynomials of degree at most $\ell - 1$. Then there exists a nonsingular matrix $P \otimes I$ such that $\mathbf{V}_\ell^{\tilde{p}}(X, \Lambda) = (P \otimes I) \cdot \mathbf{V}_\ell^p(X, \Lambda)$ for any pair (X, Λ), where both $\mathbf{V}_\ell^{\tilde{p}}(X, \Lambda)$ and $\mathbf{V}_\ell^p(X, \Lambda)$ are defined as in* (3.10).

Proof. Let

$$\tilde{p}_i(\lambda) = p_{i,0} \cdot p_0(\lambda) + \cdots + p_{i,\ell-1} \cdot p_{\ell-1}(\lambda), \qquad i = 0, \dots, \ell - 1$$

be the expansion of the polynomials $\tilde{p}_0, \dots, \tilde{p}_\ell$ in the polynomial basis formed by $p_0, \dots, p_{\ell-1}$. Since both sets of polynomials constitute bases, the matrix

$$P = \begin{bmatrix} p_{0,0} & \cdots & p_{0,\ell-1} \\ \vdots & \ddots & \vdots \\ p_{\ell-1,0} & \cdots & p_{\ell-1,\ell-1} \end{bmatrix},$$

and hence also, $P \otimes I$, is nonsingular. Furthermore, the claimed relationship $\mathbf{V}_\ell^{\tilde{p}}(X, \Lambda) = (P \otimes I) \cdot \mathbf{V}_\ell^{p}(X, \Lambda)$ holds by construction. □

Proposition 3.2.1 enables us to generalize results involving $\mathbf{V}_\ell^{p}(X, \Lambda)$ for a specific polynomial basis $p_0, \dots, p_{\ell-1}$ to arbitrary polynomial bases. This is especially useful for statements which are a lot easier to prove for a certain polynomial basis than for the others. Applying this technique to Proposition 3.1.4 confirms that the matrix $\mathbf{V}_\ell^{p}(X, \Lambda)$ can be employed equally well to characterize minimality.

Corollary 3.2.2. *Let the polynomials $p_0, \dots, p_{\ell-1}$ constitute a basis for the vector space of polynomials of degree at most $\ell - 1$. Then a pair $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ is minimal of index at most ℓ if and only if the matrix $\mathbf{V}_\ell^{p}(X, \Lambda)$ as defined in* (3.10) *has full column rank.*

Proof. Let $\tilde{p}_i(\lambda) = \lambda^i$, $i = 0, \dots, \ell - 1$. Proposition 3.1.4 and Definition 3.1.5 imply that the pair (X, Λ) is minimal of index at most ℓ if and only if the matrix $\mathbf{V}_\ell^{\tilde{p}}(X, \Lambda)$ has full column rank. The proof is now finished by inferring from Proposition 3.2.1 that $\mathbf{V}_\ell^{\tilde{p}}(X, \Lambda)$ and $\mathbf{V}_\ell^{p}(X, \Lambda)$ have the same rank. □

In the previous section, we have used a certain kind of similarity transformation to extract spectral information from a minimal invariant pair. [83, Lemma 4 (1.)] reveals how the block residual $\mathbf{T}(X, \Lambda)$ and the matrix $\mathbf{V}_\ell(X, \Lambda)$ change when the pair (X, Λ) is subjected to such a transformation. This result may be extended to cover also $\mathbf{V}_\ell^{p}(X, \Lambda)$.

Proposition 3.2.3. *Let $p_0, \dots, p_{\ell-1}$ constitute a basis for the vector space of polynomials of degree at most $\ell - 1$ and $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$. Then for any invertible matrix G,*

$$\mathbf{T}(XG, G^{-1}\Lambda G) = \mathbf{T}(X, \Lambda)G \qquad \text{and} \qquad \mathbf{V}_\ell^{p}(XG, G^{-1}\Lambda G) = \mathbf{V}_\ell^{p}(X, \Lambda)G.$$

In particular, if (X, Λ) is invariant and/or minimal, then so is $(XG, G^{-1}\Lambda G)$.

Proof. The first formula is immediately clear from the definition (3.4) of the block residual $\mathbf{T}$. The second formula is a consequence of the well-known fact that $p_i(G^{-1}\Lambda G) = G^{-1}p_i(\Lambda)G$, $i = 0, \dots, \ell - 1$. From these formulas, it is obvious that $\mathbf{T}(X, \Lambda) = 0$ implies $\mathbf{T}(XG, G^{-1}\Lambda G) = 0$ and $\mathbf{V}_\ell^{p}(XG, G^{-1}\Lambda G)$ has full column rank if and only if the same is true of $\mathbf{V}_\ell^{p}(X, \Lambda)$, thereby proving the last statement. □

3.3 Invariance and minimality of composite pairs

This section deals with minimal invariant pairs which are composed of smaller blocks. For instance, in Theorem 3.1.13, a pair of the form

$$(X, \Lambda) = \left(\begin{bmatrix} X_1, \dots, X_k \end{bmatrix}, \begin{bmatrix} \Lambda_1 & & \\ & \ddots & \\ & & \Lambda_k \end{bmatrix} \right) \tag{3.11}$$

is considered, where $X_1, \dots, X_k$ represent a set of Jordan chains and $\Lambda_1, \dots, \Lambda_k$ are the corresponding Jordan blocks. Since in this case, the Λ-component of the pair is block diagonal, a simple link can be established between the invariance and minimality of the full pair and those of its constituents $(X_1, \Lambda_1), \dots, (X_k, \Lambda_k)$.

Proposition 3.3.1. *Let the pair* (X, Λ) *be of the form* (3.11) *such that the sizes of* $X_1, \dots, X_k$ *and* $\Lambda_1, \dots, \Lambda_k$ *are compatible. Then,*

$$\mathbf{T}(X, \Lambda) = [\mathbf{T}(X_1, \Lambda_1), \dots, \mathbf{T}(X_k, \Lambda_k)],$$

and for any positive integer ℓ *as well as any polynomial basis* $p_0, \dots, p_{\ell-1}$*,*

$$\mathbf{V}_\ell^p(X, \Lambda) = [\mathbf{V}_\ell^p(X_1, \Lambda_1), \dots, \mathbf{V}_\ell^p(X_k, \Lambda_k)].$$

In particular, (X, Λ) *is invariant if and only if the same is true for each of the pairs* $(X_1, \Lambda_1), \dots, (X_k, \Lambda_k)$*. Additionally, if* (X, Λ) *is minimal, its minimality index is not smaller than the maximum of the minimality indices of* $(X_1, \Lambda_1), \dots, (X_k, \Lambda_k)$*.*

Proof. The first formula follows by straightforward calculation from the contour integral definition of the block residual in Equation (3.4). The second formula is obtained from the definition of $\mathbf{V}_\ell^p(X, \Lambda)$ in Equation (3.10) and the well-known fact (see, e.g., [60, Theorem 1.13 (g)]) that

$$p_i\bigl(\operatorname{diag}\{\Lambda_1, \dots, \Lambda_k\}\bigr) = \operatorname{diag}\bigl\{p_i(\Lambda_1), \dots, p_i(\Lambda_k)\bigr\}.$$

It is clear from the first formula that $\mathbf{T}(X, \Lambda)$ vanishes if and only if $\mathbf{T}(X_i, \Lambda_i)$ vanishes for all $i = 1, \dots, k$, implying the statement about the invariance. The second formula shows that $\mathbf{V}_\ell^p(X, \Lambda)$ cannot have full column rank unless all of $\mathbf{V}_\ell^p(X_1, \Lambda_1), \dots, \mathbf{V}_\ell^p(X_k, \Lambda_k)$ have full column rank, from which the statement about the minimality index can be concluded. □

Whereas, by Proposition 3.3.1, the invariance of the pairs $(X_1, \Lambda_1), \dots, (X_k, \Lambda_k)$ warrants the invariance of the composite pair (3.11), an analogous result does not hold, in general, for minimality; take $k = 2$, $X_1 = X_2$ and $\Lambda_1 = \Lambda_2$ for a simple counterexample. Such a statement becomes possible if we additionally require that the spectra of $\Lambda_1, \dots, \Lambda_k$ be mutually disjoint. This condition, however, is only sufficient but not necessary.

Proposition 3.3.2. *Let* $(X_1, \Lambda_1), \dots, (X_k, \Lambda_k)$ *be minimal pairs such that for any* $i, j \in \{1, \dots, k\}$ *with* $i \neq j$*,* Λ_i *and* Λ_j *have no eigenvalues in common. Then, the composite pair in* (3.11) *is minimal.*

Proof. We will prove the result for $k = 2$. The general statement then follows by an easy induction, which is omitted. Let $\left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right]$ be a solution of

$$\begin{bmatrix} X_1 & X_2 \\ \Lambda_1 - \lambda I & 0 \\ 0 & \Lambda_2 - \lambda I \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = 0$$

for some eigenvalue λ of $\left[\begin{smallmatrix} \Lambda_1 & \\ & \Lambda_2 \end{smallmatrix}\right]$. To establish the minimality of the composite pair (3.11), we need to show $\left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = 0$. W.l.o.g., assume that λ is an eigenvalue of Λ_1 but not of Λ_2. In the opposite case, analogous arguments apply. Since $\Lambda_2 - \lambda I$ is nonsingular, the last block row of the above equation entails $v = 0$. Thus, we are left with

$$\begin{bmatrix} X_1 \\ \Lambda_1 - \lambda I \end{bmatrix} u = 0,$$

which only admits the solution $u = 0$ thanks to the minimality of (X_1, Λ_1). □

3.3.1 A generalization of divided differences. The situation becomes more difficult for pairs (X, Λ) where the Λ-component possesses non-zero off-diagonal blocks coupling the constituents of the pair. Corresponding results will be derived in Section 3.3.2 further down. To be able to state these results in a convenient way, it will be beneficial to introduce a generalization to the notion of a divided difference.

Definition 3.3.3. Let $\Lambda_1 \in \mathbb{C}^{m \times m}$, $\Lambda_2 \in \mathbb{C}^{k \times k}$ be arbitrary square matrices. For any function $f : \mathbb{C}^{m \times m} \to \mathbb{C}^{n \times m}$ such that the mapping $\xi \mapsto f(\xi I_m)$ is holomorphic in a neighborhood of $\operatorname{spec} \Lambda_1 \cup \operatorname{spec} \Lambda_2$, we define the *divided difference* $\mathrm{D}_{[\Lambda_1, \Lambda_2]} f$ as the linear map

$$\begin{aligned} &\mathrm{D}_{[\Lambda_1, \Lambda_2]} f : \mathbb{C}^{m \times k} \to \mathbb{C}^{n \times k}, \\ &(\mathrm{D}_{[\Lambda_1, \Lambda_2]} f) V = \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{C}} f(\xi I_m)(\xi I_m - \Lambda_1)^{-1} V (\xi I_k - \Lambda_2)^{-1} \, \mathrm{d}\xi, \end{aligned}$$

where $\mathcal{C}$ is a contour enclosing the eigenvalues of both Λ_1 and Λ_2 in its interior.

The connection between Definition 3.3.3 and the usual notion of a divided difference is disclosed if we set $m = k = 1$. In this event, one easily calculates that

$$(\mathrm{D}_{[\Lambda_1, \Lambda_2]} f) V = \begin{cases} \frac{f(\Lambda_2) - f(\Lambda_1)}{\Lambda_2 - \Lambda_1} \cdot V, & \Lambda_1 \neq \Lambda_2, \\ \dot{f}(\Lambda_1) \cdot V, & \Lambda_1 = \Lambda_2. \end{cases}$$

For larger m and k, the division turns into the Sylvester equation

$$\Lambda_1 F - F \Lambda_2 = f(\Lambda_1) V - V f(\Lambda_2),$$

of which $F = (\mathrm{D}_{[\Lambda_1, \Lambda_2]} f) V$ is a solution. Moreover, if $\Lambda_1 = \Lambda_2$ (implying $m = k$), then $\mathrm{D}_{[\Lambda_1, \Lambda_2]} f$ coincides with the Fréchet derivative $\mathrm{D} f(\Lambda_1)$.

The generalized divided differences facilitate a more compact reformulation of a result from [77, Lemma 1.1] concerning matrix functions of composite matrices, which is summarized by the next Proposition.

Proposition 3.3.4 ([77, Lemma 1.1]). *Let $f : \mathcal{D} \to \mathbb{C}$ be holomorphic on an open set $\mathcal{D} \subset \mathbb{C}$ and denote the corresponding matrix function by f as well. Then,*

$$f\left(\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix}\right) = \begin{bmatrix} f(\Lambda_{11}) & (\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}f)\Lambda_{12} \\ 0 & f(\Lambda_{22}) \end{bmatrix},$$

provided that Λ_{11} and Λ_{22} are square matrices with $\operatorname{spec}\Lambda_{11} \cup \operatorname{spec}\Lambda_{22} \subset \mathcal{D}$.

Proof. By the contour integral representation of matrix functions, we have

$$f\left(\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix}\right) = \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{C}} f(\xi) \begin{bmatrix} \xi I - \Lambda_{11} & -\Lambda_{12} \\ 0 & \xi I - \Lambda_{22} \end{bmatrix}^{-1} \mathrm{d}\xi,$$

where $\mathcal{C}$ is a contour in $\mathcal{D}$ enclosing the eigenvalues of both Λ_{11} and Λ_{22} in its interior. The claim is now established by exploiting

$$\begin{bmatrix} \xi I - \Lambda_{11} & -\Lambda_{12} \\ 0 & \xi I - \Lambda_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (\xi I - \Lambda_{11})^{-1} & (\xi I - \Lambda_{11})^{-1}\Lambda_{12}(\xi I - \Lambda_{22})^{-1} \\ 0 & (\xi I - \Lambda_{22})^{-1} \end{bmatrix} \tag{3.12}$$

together with Definition 3.3.3. □

Besides matrix functions, we will also consider divided differences for the functions

$$\mathbf{T}(X,\cdot) : \mathbb{C}^{m\times m} \to \mathbb{C}^{n\times m}, \quad \Lambda \mapsto \mathbf{T}(X,\Lambda)$$

and

$$\mathbf{V}_\ell^p(X,\cdot) : \mathbb{C}^{m\times m} \to \mathbb{C}^{\ell n\times m}, \quad \Lambda \mapsto \mathbf{V}_\ell^p(X,\Lambda).$$

Thanks to the special structure of these functions, the expressions for their corresponding divided differences can be simplified.

Lemma 3.3.5. *The divided difference corresponding to the block residual is given by*

$$\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}\mathbf{T}(X,\cdot)\Lambda_{12} = \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{C}} T(\xi)X(\xi I - \Lambda_{11})^{-1}\Lambda_{12}(\xi I - \Lambda_{22})^{-1}\,\mathrm{d}\xi, \tag{3.13}$$

where $\mathcal{C}$ is a contour enclosing the eigenvalues of both Λ_{11} and Λ_{22} in its interior. If the underlying matrix-valued function T is of the special form (3.1)*, then the above formula is equivalent to*

$$\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}\mathbf{T}(X,\cdot)\Lambda_{12} = T_1X(\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}f_1)\Lambda_{12} + \cdots + T_dX(\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}f_d)\Lambda_{12}.$$

Proof. From the contour integral definition (3.4) of the block residual, it is easily seen that $\mathbf{T}(X,\xi I) = T(\xi)X$, which together with Definition 3.3.3 results in (3.13). The second formula can be deduced from (3.13) by inserting the special form (3.1) of the matrix-valued function and utilizing the linearity of the contour integral. □

Lemma 3.3.6. *For any positive integer ℓ and any polynomial basis $p_0,\dots,p_{\ell-1}$,*

$$\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}\mathbf{V}_\ell^p(X,\cdot)\Lambda_{12} = \begin{bmatrix} X(\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}p_0)\Lambda_{12} \\ \vdots \\ X(\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}p_{\ell-1})\Lambda_{12} \end{bmatrix}. \tag{3.14}$$

Proof. Let $\mathcal{C}$ be a contour enclosing the eigenvalues of both Λ_{11} and Λ_{22} in its interior. From the definition of $\mathbf{V}_\ell^p(X,\Lambda)$ in (3.10), we conclude that $\mathbf{V}_\ell^p(X,\xi I) = [p_0(\xi)X^\mathsf{T},\dots,p_{\ell-1}(\xi)X^\mathsf{T}]^\mathsf{T}$. Thus,

$$\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}\mathbf{V}_\ell^p(X,\cdot)\Lambda_{12} = \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}} \begin{bmatrix} p_0(\xi)X \\ \vdots \\ p_{\ell-1}(\xi)X \end{bmatrix} (\xi I - \Lambda_{11})^{-1}\Lambda_{12}(\xi I - \Lambda_{22})^{-1}\,\mathrm{d}\xi.$$

Rearranging the contour integral on the right-hand side finally yields (3.14). □

3.3.2 Composite pairs with coupling. We are now in the position to state our main results on composite pairs, which will be required for the developments in Chapters 4 and 6 but may as well be of independent interest.

Proposition 3.3.7. *Let* $\mathbf{T}(X,\Lambda)$ *and* $\mathbf{V}_\ell^p(X,\Lambda)$ *be defined as in* (3.4) *and* (3.10), *respectively, where*

$$(X,\Lambda) = \left([X_1,X_2],\ \begin{bmatrix}\Lambda_{11} & \Lambda_{12}\\ 0 & \Lambda_{22}\end{bmatrix}\right),$$

ℓ *is any positive integer, and* $p_0,\dots,p_{\ell-1}$ *constitute a basis for the vector space of polynomials of degree at most* $\ell-1$. *Then,*

$$\mathbf{T}(X,\Lambda) = \Big[\mathbf{T}(X_1,\Lambda_{11}),\ \mathbf{T}(X_2,\Lambda_{22}) + \mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}\mathbf{T}(X_1,\cdot)\Lambda_{12}\Big]$$

and

$$\mathbf{V}_\ell^p(X,\Lambda) = \Big[\mathbf{V}_\ell^p(X_1,\Lambda_{11}),\ \mathbf{V}_\ell^p(X_2,\Lambda_{22}) + \mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}\mathbf{V}_\ell^p(X_1,\cdot)\Lambda_{12}\Big]$$

with $\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}\mathbf{T}(X,\cdot)\Lambda_{12}$ *and* $\mathrm{D}_{[\Lambda_{11},\Lambda_{22}]}\mathbf{V}_\ell^p(X,\cdot)\Lambda_{12}$ *as given in* (3.13) *and* (3.14), *provided that* Λ_{11} *and* Λ_{22} *are square matrices whose sizes fit the column dimensions of* X_1 *and* X_2, *respectively.*

Proof. Let $\mathcal{C}$ be a contour enclosing the eigenvalues of both Λ_{11} and Λ_{22} in its interior. Then, by definition of the block residual,

$$\mathbf{T}\left([X_1,X_2],\begin{bmatrix}\Lambda_{11} & \Lambda_{12}\\ 0 & \Lambda_{22}\end{bmatrix}\right) = \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}} T(\xi)[X_1,X_2]\begin{bmatrix}\xi I - \Lambda_{11} & -\Lambda_{12}\\ 0 & \xi I - \Lambda_{22}\end{bmatrix}^{-1}\mathrm{d}\xi.$$

The first formula is established by inverting the block matrix as in Equation (3.12). Furthermore, for $i = 0,\dots,\ell-1$, the blocks of the matrix $\mathbf{V}_\ell^p\big([X_1,X_2],\left[\begin{smallmatrix}\Lambda_{11} & \Lambda_{12}\\ 0 & \Lambda_{22}\end{smallmatrix}\right]\big)$ are given by

$$[X_1,X_2]\cdot p_i\left(\begin{bmatrix}\Lambda_{11} & \Lambda_{12}\\ 0 & \Lambda_{22}\end{bmatrix}\right),$$

which, by Proposition 3.3.4, confirms the second formula. □

For the sake of simplicity, Proposition 3.3.7 only deals with pairs (X,Λ) where X is made up of two blocks and Λ is a 2×2 block matrix. Pairs consisting of more blocks can be handled by applying the proposition repeatedly.

The statement of Proposition 3.3.2 about minimality of composite pairs can also be generalized to the case where coupling is present; see Lemma 6.2.3 for an example.

3.3.3 Derivatives of block residuals. As is obvious from its definition in Equation (3.4), the block residual $\mathbf{T}(X, \Lambda)$ is linear in the first argument. Hence, its Fréchet derivative is given by

$$\mathrm{D}\mathbf{T}(X, \Lambda)(\triangle X, \triangle\Lambda) = \mathbf{T}(\triangle X, \Lambda) + \mathrm{D}_\Lambda \mathbf{T}(X, \Lambda)\triangle\Lambda. \tag{3.15}$$

Since $\mathrm{D}_\Lambda \mathbf{T}(X, \Lambda)\triangle\Lambda$ is again linear in X, formulas for higher-order derivatives can be obtained recursively.

Lemma 3.3.8. *For any positive integer k,*

$$\mathrm{D}^k \mathbf{T}(X, \Lambda)(\triangle X, \triangle\Lambda)^k = k \cdot \mathrm{D}_\Lambda^{k-1} \mathbf{T}(\triangle X, \Lambda)\triangle\Lambda^{k-1} + \mathrm{D}_\Lambda^k \mathbf{T}(X, \Lambda)\triangle\Lambda^k,$$

where

$$\mathrm{D}_\Lambda^k \mathbf{T}(X, \Lambda)\triangle\Lambda^k = \frac{k!}{2\pi\mathrm{i}} \int_\Gamma T(\xi)X(\xi I - \Lambda)^{-1}\left[\triangle\Lambda(\xi I - \Lambda)^{-1}\right]^k \mathrm{d}\xi$$

with a contour Γ enclosing the eigenvalues of Λ in its interior.

Proof. We begin by demonstrating the second formula. For sufficiently small $\|\triangle\Lambda\|$, a Neumann series argument yields

$$\begin{aligned}\left[\xi I - (\Lambda + \triangle\Lambda)\right]^{-1} &= (\xi I - \Lambda)^{-1}\left[I - \triangle\Lambda(\xi I - \Lambda)^{-1}\right]^{-1} \\ &= (\xi I - \Lambda)^{-1} \sum_{k=0}^{\infty} \left[\triangle\Lambda(\xi I - \Lambda)^{-1}\right]^k,\end{aligned}$$

showing that $\mathrm{D}_\Lambda^k (\xi I - \Lambda)^{-1}\triangle\Lambda^k = k! \cdot (\xi I - \Lambda)^{-1}\left[\triangle\Lambda(\xi I - \Lambda)^{-1}\right]^k$. Furthermore, from the contour integral definition (3.4) of the block residual, we find

$$\mathrm{D}_\Lambda^k \mathbf{T}(X, \Lambda)\triangle\Lambda^k = \frac{1}{2\pi\mathrm{i}} \int_\Gamma T(\xi)X \cdot \mathrm{D}_\Lambda^k (\xi I - \Lambda)^{-1}\triangle\Lambda^k \,\mathrm{d}\xi$$

with Γ as described above. Exchanging the differentiation with the contour integral is feasible whenever the integral on the right-hand side exists. The latter is proven along with the second identity of the lemma by combining the last two results.

The proof of the first identity is by induction. By the discussion immediately preceding this lemma, the identity holds true for $k = 1$. So assume the identity is valid for some k. Since, by the second formula of the lemma, $\mathrm{D}_\Lambda^k \mathbf{T}(X, \Lambda)\triangle\Lambda^k$ is linear in X, we have

$$\begin{aligned}&\mathrm{D}^{k+1}\mathbf{T}(X, \Lambda)(\triangle X, \triangle\Lambda)^{k+1} \\ &= \mathrm{D}_{(X,\Lambda)}\left[k \cdot \mathrm{D}_\Lambda^{k-1}\mathbf{T}(\triangle X, \Lambda)\triangle\Lambda^{k-1} + \mathrm{D}_\Lambda^k \mathbf{T}(X, \Lambda)\triangle\Lambda^k\right](\triangle X, \triangle\Lambda) \\ &= (k+1) \cdot \mathrm{D}_\Lambda^k \mathbf{T}(\triangle X, \Lambda)\triangle\Lambda^k + \mathrm{D}_\Lambda^{k+1}\mathbf{T}(X, \Lambda)\triangle\Lambda^{k+1}\end{aligned}$$

Thus, the identity holds true also for $k + 1$ and the proof is complete. □

Comparing the second formula of Lemma 3.3.8 for $k = 1$ with Equation (3.13) shows that the particular divided difference $\mathrm{D}_{[\Lambda,\Lambda]}\mathbf{T}(X, \cdot)\triangle\Lambda$ coincides with the derivative $\mathrm{D}_\Lambda \mathbf{T}(X, \Lambda)\triangle\Lambda$. As a consequence, it is possible to draw connections between the derivative(s) of a block residual and the block residual of a composite pair, similar to the results for matrix functions in [94]. In the following, we give two examples.

Proposition 3.3.9. *Let the block residual be defined as in Equation* (3.4). *Then, for any $X, \triangle X \in \mathbb{C}^{n\times m}$ and $\Lambda, \triangle\Lambda \in \mathbb{C}^{m\times m}$,*

$$\mathbf{T}\left([X, \triangle X], \begin{bmatrix} \Lambda & \triangle\Lambda \\ 0 & \Lambda \end{bmatrix}\right) = \big[\mathbf{T}(X,\Lambda),\ \mathrm{D}\mathbf{T}(X,\Lambda)(\triangle X, \triangle\Lambda)\big].$$

If, additionally, $Y \in \mathbb{C}^{n\times m}$ and $\Theta \in \mathbb{C}^{m\times m}$, then

$$\mathbf{T}\left(\big[X, \triangle X, -\tfrac{1}{2}Y\big], \begin{bmatrix} \Lambda & \triangle\Lambda & -\frac{1}{2}\Theta \\ & \Lambda & \triangle\Lambda \\ & & \Lambda \end{bmatrix}\right)$$
$$= \Big[\mathbf{T}(X,\Lambda),\ \mathrm{D}\mathbf{T}(X,\Lambda)(\triangle X, \triangle\Lambda),\ \tfrac{1}{2}\big(\mathrm{D}^2\mathbf{T}(X,\Lambda)(\triangle X, \triangle\Lambda)^2 - \mathrm{D}\mathbf{T}(X,\Lambda)(Y,\Theta)\big)\Big].$$

Proof. From Proposition 3.3.7, we infer

$$\mathbf{T}\left([X, \triangle X], \begin{bmatrix} \Lambda & \triangle\Lambda \\ 0 & \Lambda \end{bmatrix}\right) = \big[\mathbf{T}(X,\Lambda),\ \mathbf{T}(\triangle X,\Lambda) + \mathrm{D}_{[\Lambda,\Lambda]}\mathbf{T}(X,\cdot)\triangle\Lambda\big].$$

As $\mathrm{D}_{[\Lambda,\Lambda]}\mathbf{T}(X,\cdot)\triangle\Lambda = \mathrm{D}_\Lambda\mathbf{T}(X,\Lambda)\triangle\Lambda$ by the discussion preceding this proposition, the second block on the right-hand side amounts to $\mathrm{D}\mathbf{T}(X,\Lambda)(\triangle X, \triangle\Lambda)$ according to Equation (3.15). This concludes the proof of the first identity.

We now turn to the second identity. Splitting the pair after the second block, Proposition 3.3.7 implies

$$\mathbf{T}\left(\big[X, \triangle X, -\tfrac{1}{2}Y\big], \begin{bmatrix} \Lambda & \triangle\Lambda & -\frac{1}{2}\Theta \\ & \Lambda & \triangle\Lambda \\ & & \Lambda \end{bmatrix}\right) = \big[\mathbf{T}_1, \mathbf{T}_2\big],$$

where the first block on the right-hand side,

$$\mathbf{T}_1 = \mathbf{T}\left([X, \triangle X], \begin{bmatrix} \Lambda & \triangle\Lambda \\ 0 & \Lambda \end{bmatrix}\right),$$

already has the desired structure by the first identity, and the second block reads

$$\mathbf{T}_2 = \mathbf{T}(-\tfrac{1}{2}Y,\Lambda) + \frac{1}{2\pi\mathrm{i}}\int_\Gamma T(\xi)[X,\triangle X]\begin{bmatrix} \xi I - \Lambda & -\triangle\Lambda \\ 0 & \xi I - \Lambda \end{bmatrix}^{-1}\begin{bmatrix} -\frac{1}{2}\Theta \\ \triangle\Lambda \end{bmatrix}(\xi I - \Lambda)^{-1}\,\mathrm{d}\xi.$$

Here, the contour Γ needs to be chosen such that it encloses all eigenvalues of Λ. By inverting the block matrix and using Lemma 3.3.8 as well as the linearity in the first argument of the block residual, the second block simplifies to

$$\begin{aligned}\mathbf{T}_2 &= -\tfrac{1}{2}\mathbf{T}(Y,\Lambda) - \tfrac{1}{2}\mathrm{D}_\Lambda\mathbf{T}(X,\Lambda)\Theta + \tfrac{1}{2}\mathrm{D}_\Lambda^2\mathbf{T}(X,\Lambda)\triangle\Lambda^2 + \mathrm{D}_\Lambda\mathbf{T}(\triangle X,\Lambda)\triangle\Lambda \\ &= \tfrac{1}{2}\big(\mathrm{D}^2\mathbf{T}(X,\Lambda)(\triangle X,\triangle\Lambda)^2 - \mathrm{D}\mathbf{T}(X,\Lambda)(Y,\Theta)\big).\end{aligned}$$

Thus, also the second identity is proved. □

The occurrence of block triangular block Toeplitz matrices in the second argument of the block residual, which can be observed in Proposition 3.3.9, is quite characteristic for this sort of results. For the next statement, it will therefore be convenient to adopt the shorthand notation (cf. [94, Section 4])

$$\mathrm{BTT}_k[B_1, \dots, B_k] = \begin{bmatrix} B_1 & B_2 & \cdots & B_k \\ & B_1 & \ddots & \vdots \\ & & \ddots & B_2 \\ & & & B_1 \end{bmatrix} \in \mathbb{C}^{km \times km}$$

for the block upper triangular block Toeplitz matrix composed of the square matrix blocks $B_1, \dots, B_k \in \mathbb{C}^{m \times m}$.

Proposition 3.3.10. *Let the block residual be defined as in Equation* (3.4) *and let* $X, \triangle X \in \mathbb{C}^{n \times m}$, $\Lambda, \triangle\Lambda \in \mathbb{C}^{m \times m}$. *For any positive integer* k, *define*

$$X_k = [X, \triangle X, \underbrace{0, \dots, 0}_{k-1 \text{ times}}] \in \mathbb{C}^{n \times (k+1)m}, \quad \Lambda_k = \mathrm{BTT}_{k+1}[\Lambda, \triangle\Lambda, \underbrace{0, \dots, 0}_{k-1 \text{ times}}],$$

where the zero blocks in X_k *are of size* $n \times m$ *and the zero blocks in* Λ_k *of size* $m \times m$. *Then,*

$$\mathbf{T}(X_k, \Lambda_k) = [\mathbf{T}(X, \Lambda),\ \mathrm{D}\mathbf{T}(X, \Lambda)(\triangle X, \triangle\Lambda), \dots, \tfrac{1}{k!}\mathrm{D}^k\mathbf{T}(X, \Lambda)(\triangle X, \triangle\Lambda)^k].$$

Proof. Let Γ be a contour enclosing the eigenvalues of Λ in its interior. Then, by the contour integral definition (3.4) of the block residual,

$$\mathbf{T}(X_k, \Lambda_k) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma T(\xi) X_k (\xi I - \Lambda_k)^{-1} \,\mathrm{d}\xi.$$

Block matrix inversion of the block upper triangular matrix $(\xi I - \Lambda_k)^{-1}$ and subsequent block matrix multiplication by X_k shows that

$$\mathbf{T}(X_k, \Lambda_k) = \big[\mathbf{T}_0, \dots, \mathbf{T}_k\big],$$

where $\mathbf{T}_0 = \dfrac{1}{2\pi \mathrm{i}} \displaystyle\int_\Gamma T(\xi) X (\xi I - \Lambda)^{-1} \,\mathrm{d}\xi$ and, for $j = 1, \dots, k$,

$$\mathbf{T}_j = \frac{1}{2\pi \mathrm{i}} \int_\Gamma T(\xi) \big[X(\xi I - \Lambda)^{-1} \triangle\Lambda + \triangle X\big] (\xi I - \Lambda)^{-1} \big[\triangle\Lambda (\xi I - \Lambda)^{-1}\big]^{j-1} \,\mathrm{d}\xi.$$

Clearly, $\mathbf{T}_0 = \mathbf{T}(X, \Lambda)$. Furthermore, Lemma 3.3.8 shows that

$$\mathbf{T}_j = \tfrac{1}{j!}\mathrm{D}^j\mathbf{T}(X, \Lambda)(\triangle X, \triangle\Lambda)^j, \quad j = 1, \dots, k,$$

thereby concluding the proof. □

As a major consequence, Proposition 3.3.10 implies that derivatives of a block residual can be conveniently determined by evaluating the block residual of an appropriately augmented pair. Exploiting that $\mathrm{D}^k\mathbf{T}(I, \lambda I)(0, I)^k = \frac{\mathrm{d}^k}{\mathrm{d}\lambda^k} T(\lambda)$ for any scalar λ by Lemma 3.3.8, we can, in particular, extract derivatives of any order k for the underlying matrix-valued function T by computing the block residual

$$\mathbf{T}\big([I, 0, 0, \dots, 0], \mathrm{BTT}_{k+1}[\lambda I, I, 0, \dots, 0]\big) = \big[T(\lambda), \dot{T}(\lambda), \dots, \tfrac{1}{k!}\tfrac{\mathrm{d}^k}{\mathrm{d}\lambda^k} T(\lambda)\big].$$

Thanks to the above facts, the block residual offers an attractive alternative for designing the interface to a nonlinear eigensolver. The user specifies the nonlinear eigenvalue problem by supplying a subroutine which computes the block residual for an arbitrary pair. Any additional problem-related information the solver might need is then deduced from block residuals of specially structured pairs. This way, the user can apply problem-specific techniques for computing the block residuals, which would otherwise not be possible in a general-purpose method. This gain is especially profitable for solvers which inherently require the formation of block residuals, such as the algorithms developed in Chapters 4 and 6. For this sort of algorithms, the block-residual approach is clearly superior to the conventional method of supplying a subroutine for evaluating the matrix-valued function T as well as its derivatives. Nevertheless, virtually every existing nonlinear eigensolver, including the algorithms in Section 1.2, can be redesigned to handle block residuals as user input.

3.4 Extraction and embedding of minimal invariant pairs

In [16, Theorem 3], the possibility was raised to extract a minimal invariant pair of a matrix polynomial from a non-minimal one. In the following, we will refine this result and transfer it to the setting of general nonlinear eigenvalue problems. The extraction of minimal invariant pairs will then be employed for a series of proofs later in this section as well as in Chapter 4.

Proposition 3.4.1. *Let $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ and assume that $\mathbf{V}_m(X, \Lambda)$ has rank r. Then, for any invertible matrix $S \in \mathbb{C}^{m\times m}$ such that $\mathbf{V}_m(X, \Lambda)S = [V,\ 0]$ with $V \in \mathbb{C}^{mn\times r}$, we have*

$$XS = [X_1,\ 0], \quad S^{-1}\Lambda S = \begin{bmatrix} \Lambda_{11} & 0 \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}.$$

Furthermore, $(X_1, \Lambda_{11}) \in \mathbb{C}^{n\times r} \times \mathbb{C}^{r\times r}$ is a minimal pair with $\operatorname{spec} \Lambda_{11} \subset \operatorname{spec} \Lambda$. If, additionally, (X, Λ) is invariant, then so is (X_1, Λ_{11}).

Proof. Recall that $\mathbf{V}_m(X, \Lambda)S = \mathbf{V}_m(XS, S^{-1}\Lambda S)$ by Proposition 3.2.3. Hence, the last $m - r$ columns of $XS(S^{-1}\Lambda S)^k \in C^{n\times m}$ are zero for $k = 0, \dots, m - 1$, which we denote as

$$XS(S^{-1}\Lambda S)^k = [*,\ 0]. \tag{3.16}$$

Setting $k = 0$ yields $XS = [X_1,\ 0]$ for some suitably chosen $X_1 \in \mathbb{C}^{n\times r}$. Moreover, by the Caley-Hamilton theorem, $XS(S^{-1}\Lambda S)^m$ is a linear combination of $XS(S^{-1}\Lambda S)^k$ for $k = 0, \dots, m - 1$, showing that Equation 3.16 also holds for $k = m$. Partitioning

$$S^{-1}\Lambda S = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$$

conformally with $[V,\ 0]$, we therefore have

$$[*,\ 0] = \mathbf{V}_m(XS, S^{-1}\Lambda S) \cdot (S^{-1}\Lambda S) = [V,\ 0] \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} = [V\Lambda_{11},\ V\Lambda_{12}].$$

Since S is invertible, V has full column rank. Therefore, the last block column of the above equation implies $\Lambda_{12} = 0$. Exploiting the structure of XS and $S^{-1}\Lambda S$, we find $\mathbf{V}_m(XS, S^{-1}\Lambda S) = [\mathbf{V}_m(X_1, \Lambda_{11}),\ 0]$. Thus, $\mathbf{V}_m(X_1, \Lambda_{11}) = V$ has full column rank, showing that (X_1, Λ_{11}) is minimal.

If (X, Λ) is invariant, the same is true for $(XS, S^{-1}\Lambda S)$ by Proposition 3.2.3. Again using the structure of XS and $S^{-1}\Lambda S$, the contour integral definition 3.4 of the block residual gives

$$\mathbf{T}(XS, S^{-1}\Lambda S) = \frac{1}{2\pi\mathrm{i}} \int_\Gamma T(\xi)[X_1,\ 0] \begin{bmatrix} \xi I - \Lambda_{11} & 0 \\ -\Lambda_{21} & \xi I - \Lambda_{22} \end{bmatrix}^{-1} \mathrm{d}\xi,$$

where Γ is a contour enclosing the eigenvalues of Λ in its interior. Inverting the block matrix leads to $\mathbf{T}(XS, S^{-1}\Lambda S) = \big[\mathbf{T}(X_1, \Lambda_{11}),\ 0\big]$, which confirms the last statement. □

In the discussion preceding Definition 3.1.14, we have observed that a simple invariant pair (X, Λ) contains a full canonical system of (generalized) eigenvectors for each eigenvalue of Λ, whereas a multiple invariant pair holds only part of this information. This observation indicates that it should be possible to embed a given minimal invariant pair into a simple one comprising the same eigenvalues. The subsequent proposition realizes this idea. In fact, the assumption of minimality can even be weakened.

Proposition 3.4.2. *Let $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ be an invariant, not necessarily minimal pair of some matrix-valued function T and $(\hat{X}, \hat{\Lambda}) \in \mathbb{C}^{n\times \hat{m}} \times \mathbb{C}^{\hat{m}\times \hat{m}}$ a simple invariant pair of T such that $\operatorname{spec}\Lambda \subseteq \operatorname{spec}\hat{\Lambda}$. Then there exists a unique matrix $C \in \mathbb{C}^{\hat{m}\times m}$ satisfying*

$$\hat{X}C = X, \qquad \hat{\Lambda}C = C\Lambda.$$

Furthermore, if (X, Λ) is minimal, C has full column rank and for any invertible matrix $\hat{C} = [C, C_1] \in \mathbb{C}^{\hat{m}\times \hat{m}}$,

$$(\hat{X}\hat{C},\ \hat{C}^{-1}\hat{\Lambda}\hat{C}) = \left([X, *],\ \begin{bmatrix} \Lambda & * \\ 0 & * \end{bmatrix}\right),$$

where $*$ signifies an arbitrary matrix block of appropriate size.

Proof. By Proposition 3.3.1, the pair

$$(X_*,\ \Lambda_*) := \left([\hat{X}, X],\ \begin{bmatrix} \hat{\Lambda} & \\ & \Lambda \end{bmatrix}\right)$$

is invariant, and the rank of $\mathbf{V}_{\hat{m}+1}(X_*, \Lambda_*) = \big[\mathbf{V}_{\hat{m}+1}(\hat{X}, \hat{\Lambda}),\ \mathbf{V}_{\hat{m}+1}(X, \Lambda)\big]$ is at least $\hat{m}$. On the other hand, since $\operatorname{spec}\Lambda_* = \operatorname{spec}\hat{\Lambda}$, the rank cannot exceed $\hat{m}$. Otherwise, by Proposition 3.4.1, a minimal invariant pair of size larger than $\hat{m}$ could be extracted from (X_*, Λ_*), contradicting the fact that $(\hat{X}, \hat{\Lambda})$ is simple. Consequently, there exists a matrix $C \in \mathbb{C}^{\hat{m}\times m}$ such that $\mathbf{V}_{\hat{m}+1}(X, \Lambda) = \mathbf{V}_{\hat{m}+1}(\hat{X}, \hat{\Lambda})C$ or, recalling the definition of $\mathbf{V}_{\hat{m}+1}$ in (3.6),

$$X\Lambda^j = \hat{X}\hat{\Lambda}^j C, \qquad j = 0, \ldots, \hat{m}.$$

Because $\mathbf{V}_{\hat{m}+1}(\hat{X},\hat{\Lambda})$ has full column rank, C is unique and has the same rank as $\mathbf{V}_{\hat{m}+1}(X,\Lambda)$. Thus, C has full column rank if (X,Λ) is minimal. Setting $j=0$ in the above equation implies $X=\hat{X}C$. Furthermore, we conclude

$$\mathbf{V}_{\hat{m}}(\hat{X},\hat{\Lambda})(\hat{\Lambda}C-C\Lambda)=\begin{bmatrix}\hat{X}\hat{\Lambda}^1\\ \vdots\\ \hat{X}\hat{\Lambda}^{\hat{m}}\end{bmatrix}C-\begin{bmatrix}\hat{X}\hat{\Lambda}^0\\ \vdots\\ \hat{X}\hat{\Lambda}^{\hat{m}-1}\end{bmatrix}C\Lambda=\begin{bmatrix}X\Lambda^1\\ \vdots\\ X\Lambda^{\hat{m}}\end{bmatrix}-\begin{bmatrix}X\Lambda^0\\ \vdots\\ X\Lambda^{\hat{m}-1}\end{bmatrix}\Lambda=0,$$

entailing $\hat{\Lambda}C=C\Lambda$ since $\mathbf{V}_{\hat{m}}(\hat{X},\hat{\Lambda})$ has full column rank.

For the final statement, let $\hat{C}=[C,C_1]$ be invertible. Then by straightforward calculation,

$$\hat{X}\hat{C}=\big[\hat{X}C,\,*\big]=\big[X,\,*\big],\quad \hat{C}^{-1}\hat{\Lambda}\hat{C}=\hat{C}^{-1}\big[\hat{\Lambda}C,\,*\big]=\hat{C}^{-1}\big[C\Lambda,\,*\big]=\begin{bmatrix}\Lambda & *\\ 0 & *\end{bmatrix},$$

as claimed. □

By the same reasoning as above, it should be possible to augment a multiple invariant pair until it becomes simple. This is put into practice by our next result.

Proposition 3.4.3. *Let $(X,\Lambda)\in\mathbb{C}^{n\times m}\times\mathbb{C}^{m\times m}$ be a minimal invariant pair with multiplicity s of some matrix-valued function T. Then either $s=1$ or there exist vectors $y\in\mathbb{C}^n$, $z\in\mathbb{C}^m$ and a scalar $\theta\in\operatorname{spec}\Lambda$ such that*

$$\left(\big[X,\,y\big],\,\begin{bmatrix}\Lambda & z\\ 0 & \theta\end{bmatrix}\right) \tag{3.17}$$

constitutes a minimal invariant pair of multiplicity $s-1$.

Proof. If $s=1$, there is nothing to show. For $s>1$, let $(\hat{X},\hat{\Lambda})$ be a simple invariant pair of T with $\operatorname{spec}\hat{\Lambda}=\operatorname{spec}\Lambda$. By Proposition 3.4.2, there exists an invertible matrix $\hat{C}$ with

$$\big(\hat{X}\hat{C},\,\hat{C}^{-1}\hat{\Lambda}\hat{C}\big)=\left(\big[X,\,Y\big],\,\begin{bmatrix}\Lambda & Z\\ 0 & \Theta\end{bmatrix}\right).$$

Note that $\hat{C}$ can be chosen such that Θ is in Schur form. Moreover, $(\hat{X}\hat{C},\,\hat{C}^{-1}\hat{\Lambda}\hat{C})$ is minimal invariant by Proposition 3.2.3. Partitioning

$$Y=\big[y,\,*\big],\quad Z=\big[z,\,*\big],\quad \Theta=\begin{bmatrix}\theta & *\\ 0 & *\end{bmatrix}$$

with $y\in\mathbb{C}^n$, $z\in C^m$, and $\theta\in\mathbb{C}$, Proposition 3.3.7 implies that the augmented pair (3.17) inherits the invariance and minimality of $(\hat{X}\hat{C},\,\hat{C}^{-1}\hat{\Lambda}\hat{C})$. Furthermore, $\theta\in\operatorname{spec}\Theta\subseteq\operatorname{spec}\hat{\Lambda}=\operatorname{spec}\Lambda$ as claimed and the statement about the multiplicity of the augmented pair follows by construction. □

3.4.1 An application. The extraction and embedding mechanisms described above can be used to prove a variety of statements about nonlinear eigenvalue problems in an elegant and intuitive fashion. To illustrate this fact, we will consider a statement about the null space of nonlinear Sylvester operators. Further proofs employing these techniques can be found in Chapter 4.

Let $T : \mathcal{D} \to \mathbb{C}^{n\times n}$ be a holomorphic, matrix-valued function and $\Lambda \in \mathbb{C}^{m\times m}$ a square matrix with eigenvalues in the open set $\mathcal{D}$. Then, the nonlinear Sylvester operator $\mathbf{T}_\Lambda$ associated with T and Λ is defined as

$$\mathbf{T}_\Lambda : \mathbb{C}^{n\times m} \to \mathbb{C}^{n\times m}, \quad \mathbf{T}_\Lambda(X) := \mathbf{T}(X, \Lambda),$$

where $\mathbf{T}$ denotes the block residual corresponding to T.

The null space of such nonlinear Sylvester operators has been investigated extensively in [74]. The following proposition constitutes a special case of the main result [74, Theorem 2.6] of this work. We will prove this proposition by employing extraction of minimal invariant pairs. Note also that [74, Theorem 2.6] can be obtained from the proposition with the aid of Proposition 3.3.1.

Proposition 3.4.4. *Let r be a positive integer and $\lambda \in \mathcal{D}$. Then, the following two statements are equivalent:*

(i) $\operatorname{alg}_T \lambda \geq r$*;*

(ii) $\dim\ker \mathbf{T}_{J_r(\lambda)} \geq r$*, where $J_r(\lambda)$ denotes a Jordan block of size r belonging to λ.*

Proof. Assume that (i) holds. Then, there exist Jordan chains $\{x_0^{(j)}, \dots, x_{m_j-1}^{(j)}\}$, $j = 1, \dots, g$ of T belonging to λ such that $m_1 + \cdots + m_g = r$ and $x_0^{(1)}, \dots, x_0^{(g)}$ are linearly independent. For $j = 1, \dots, g$ and $k = 0, \dots, m_j - 1$, define

$$X_k^{(j)} := \big[0, \dots, 0, x_0^{(j)}, \dots, x_k^{(j)}\big] \in \mathbb{C}^{n\times r}.$$

Using Proposition 3.3.7 and Lemma 3.1.11, we find that

$$\mathbf{T}_{J_r(\lambda)}\big(X_k^{(j)}\big) = \Big[0,\ \mathbf{T}\big(\big[x_0^{(j)}, \dots, x_k^{(j)}\big], J_{k+1}(\lambda)\big)\Big] = 0.$$

Hence, $X_k^{(j)} \in \ker \mathbf{T}_{J_r(\lambda)}$ for all $j = 1, \dots, g$ and $k = 0, \dots, m_j - 1$. Furthermore, (X, Λ) with

$$X := \big[x_0^{(1)}, \dots, x_{m_1-1}^{(1)}, \dots, x_0^{(g)}, \dots, x_{m_g-1}^{(g)}\big], \quad \Lambda := \operatorname{diag}\big\{J_{m_1}(\lambda), \dots, J_{m_g}(\lambda)\big\}$$

constitutes a minimal invariant pair of T by Proposition 3.1.12. Consequently, the matrix $\mathbf{V}_r^p(X, \Lambda)$, with the polynomial basis $p_i(\mu) = (\mu - \lambda)^{r-1-i}$, $i = 0, \dots, r-1$, has full column rank by Corollary 3.2.2 and Corollary 3.1.7. One now readily verifies that

$$\begin{aligned} \mathbf{V}_r^p(X, \Lambda) &= \begin{bmatrix} X(\Lambda - \lambda I)^{r-1} \\ \vdots \\ X(\Lambda - \lambda I) \\ X \end{bmatrix} \\ &= \Big[\operatorname{vec} X_0^{(1)}, \dots, \operatorname{vec} X_{m_1-1}^{(1)}, \dots, \operatorname{vec} X_0^{(g)}, \dots, \operatorname{vec} X_{m_g-1}^{(g)}\Big], \end{aligned}$$

where $\operatorname{vec} X_k^{(j)}$ denotes the vectorization of the matrix $X_k^{(j)}$, i.e., the vector consisting of all columns of $X_k^{(j)}$, stacked on top of each other. We conclude that $X_k^{(j)}$ for $j = 1, \dots, g$ and $k = 0, \dots, m_j - 1$ are r linearly independent elements of $\ker \mathbf{T}_{J_r(\lambda)}$, confirming (ii).

For the converse direction, let (ii) be fulfilled and select r linearly independent elements $X_1, \ldots, X_r \in \ker \mathbf{T}_{J_r(\lambda)}$. Then the pair (X, Λ) with

$$X := [X_1, \ldots, X_r], \quad \Lambda := \operatorname{diag}\bigl\{J_r(\lambda), \ldots, J_r(\lambda)\bigr\}$$

is invariant by Proposition 3.3.1. Additionally, by a similar calculation as above, the $(k \cdot r)$-th column of

$$\mathbf{V}_r^p(X, \Lambda) = \begin{bmatrix} X(\Lambda - \lambda I)^{r-1} \\ \vdots \\ X(\Lambda - \lambda I) \\ X \end{bmatrix}$$

with $p_i(\mu) = (\mu - \lambda)^{r-1-i}$, $i = 0, \ldots, r-1$ equals $\operatorname{vec} X_k$ for $k = 1, \ldots, r$, showing that $\mathbf{V}_r^p(X, \Lambda)$ has at least rank r. Thus, by Propositions 3.4.1 and 3.1.6, a minimal invariant pair $(\hat{X}, \hat{\Lambda})$ of at least size r with $\operatorname{spec} \hat{\Lambda} = \operatorname{spec} \Lambda = \{\lambda\}$ can be extracted from (X, Λ), which entails $\operatorname{alg}_T \lambda \geq r$. □

Contributions within this chapter

In this chapter, we have reviewed and extended the theory of minimal invariant pairs. Minimal invariant pairs facilitate the numerically stable representation of several eigenpairs of a nonlinear eigenvalue problem. Therefore, the findings of this chapter form the basis for the upcoming developments in the remainder of this work.

In Section 3.1, we set the stage by defining minimal invariant pairs as well as investigating their fundamental properties. Definition 3.1.1 is more general than the original definition of invariance in [83, Definition 1] in that it prefers the block residual (3.4) based on contour integration over the original formulation (3.2). The contour integral formulation of the block residual is conceptually more elegant and does not require a special form of the matrix-valued function. To present the theory in a way deemed most intuitive by the author, Definition 3.1.3 also differs from the original definition of minimality in [83, Definition 2]. Both of the alternative definitions as well as Propositions 3.1.2 and 3.1.4 proving their equivalence to the original formulations (where applicable) have been proposed by Beyn, Kressner, and the author in [21]. The connection between the minimality index of a pair and the termination of a corresponding block Krylov subspace is unpublished work by the author. Using this connection, we also give a new proof for an existing upper bound on the minimality index in Corollary 3.1.7. Originally, this bound was shown in [83, Lemma 5]. The correspondence between minimal invariant pairs and a set of Jordan chains belonging to one (Proposition 3.1.12) or several (Theorem 3.1.13) eigenvalues of a nonlinear eigenvalue problem has been established by Beyn, Kressner, and the author in [21]. A slightly weaker version of Proposition 3.1.12 appeared before in [47, Theorem 2.3]. Finally, Definition 3.1.14 is a generalization of [22, Definition 2.1] for linear and [23, Definition 2.1] for quadratic eigenvalue problems to the general holomorphic case.

Section 3.2 is devoted to characterizing minimality using polynomial bases other than the monomials. This idea seems to be new and has been published by the author in [39]. The approach is justified by Corollary 3.2.2. Proposition 3.2.3

extends a transformation result, which has been proved for the monomial basis in [83, Lemma 4, 1.], to arbitrary polynomial bases.

In Section 3.3, we investigate pairs which are composed of smaller blocks. In particular, we examine how the block residual and the matrix $\mathbf{V}_\ell^p$ of the composite pair are related to the corresponding quantities of its building blocks. The results presented throughout Section 3.3 are new unless explicitly stated otherwise. In the beginning, pairs with a block diagonal second component are analyzed. Later, in Section 3.3.2, we cover the more general and more involved case of a block triangular second component. To enable a convenient notation of the results, we introduce a generalized notion of a divided difference in Section 3.3.1 and discuss its properties. Lastly, in Section 3.3.3, a connection between Fréchet derivatives of block residuals and block residuals of certain composite pairs is established in analogy to existing results for matrix functions in [94]. Based on these findings, it is argued that the block residual constitutes a competitive interface for software implementations of nonlinear eigensolvers.

In Section 3.4, we consider embeddings and extractions of minimal invariant pairs. Proposition 3.4.1 represents a refinement of an extraction result from [16, Theorem 3]. Propositions 3.4.2 and 3.4.3 are unpublished, new results by the author; they lend mathematical rigor to the intuition that any multiple invariant pair can be augmented until it becomes simple.

Chapter 4

Continuation of minimal invariant pairs

In applications, the matrix-valued function T in the nonlinear eigenvalue problem (1.1) often additionally depends on a number of design parameters. The goal is then to compute and track the eigenvalues of interest as these parameters vary. A prototype application serving as a motivation stems from wave propagation in a periodic medium with frequency-dependent coefficients and gives rise to a family of nonlinear eigenvalue problems depending smoothly on the wave vector. It is then of interest to compute the eigenvalues closest to the real axis for all wave vectors on the boundary and in the interior of the irreducible Brillouin zone; see also Section 2.2.

In this chapter, we will treat the special case of a nonlinear eigenvalue problem

$$T\bigl(\lambda(s), s\bigr)x(s) = 0, \quad x(s) \neq 0 \tag{4.1}$$

depending on a single real parameter s. The numerical continuation of one eigenvalue for this type of problem can be considered a classical topic in numerical analysis; see, e.g., [76, 91]. In contrast, the numerical continuation of several eigenvalues of a nonlinear eigenvalue problem has not been investigated to a large extent in the literature, with the exception of the work in [23, 16] on polynomial eigenvalue problems. Although, in principle, one could continue several eigenvalues individually, this approach bears the risk of undetected eigenvalue collisions, does not allow for eigenvalues of higher multiplicity, and can be expected to become quite challenging to implement in a robust manner. For linear eigenvalue problems, the notion of invariant subspaces offers a more convenient, elegant, and robust approach to handling several eigenvalues [22, 24, 36]. In a similar fashion, we will employ the concept of minimal invariant pairs introduced in Chapter 3 to construct a numerical continuation scheme for several eigenvalues of a nonlinear eigenvalue problem. More specifically, we will continue a minimal invariant pair of the parameter-dependent nonlinear eigenvalue problem (4.1) as the parameter s traverses a prescribed real interval. Special attention is paid to generic bifurcations encountered as eigenvalues included in the minimal invariant pair being continued coalesce with eigenvalues outside the pair.

4.1 Characterization of minimal invariant pairs

This section will provide the theoretical foundations of our continuation method. One of the major tools will be the characterization of minimal invariant pairs as solutions to certain nonlinear equations, similarly as in [83].

Let (X_0, Λ_0) be a minimal invariant pair with minimality index not exceeding ℓ. According to Definition 3.1.1, the invariance of the pair (X_0, Λ_0) is equivalent to $(X, \Lambda) = (X_0, \Lambda_0)$ being a solution of the nonlinear equation

$$\mathbf{T}(X, \Lambda) = 0.$$

Furthermore, by Proposition 3.1.4, a pair (X, Λ) is minimal with minimality index at most ℓ if and only if the matrix $\mathbf{V}_\ell(X, \Lambda)$ defined in (3.6) has full column rank. This motivates the normalization condition $\mathbf{N}(X, \Lambda) = 0$ with

$$\mathbf{N}(X, \Lambda) = W^{\mathsf{H}} \big[\mathbf{V}_\ell(X, \Lambda) - \mathbf{V}_\ell(X_0, \Lambda_0) \big], \tag{4.2}$$

where $W \in \mathbb{C}^{\ell n \times m}$ is chosen such that $W^{\mathsf{H}} \mathbf{V}_\ell(X_0, \Lambda_0)$ is invertible. Altogether, we obtain that $(X, \Lambda) = (X_0, \Lambda_0)$ satisfies the nonlinear equation

$$\mathbf{F}(X, \Lambda) = 0, \tag{4.3}$$

where

$$\mathbf{F}(X, \Lambda) = \begin{bmatrix} \mathbf{T}(X, \Lambda) \\ \mathbf{N}(X, \Lambda) \end{bmatrix}. \tag{4.4}$$

Conversely, it is easy to see that every solution (X, Λ) of (4.3) constitutes a minimal invariant pair with minimality index at most ℓ.

Since $\mathbf{F}$ is derived from $\mathbf{V}_\ell$ as well as the block residual $\mathbf{T}$, it inherits their transformation properties described in Proposition 3.2.3.

Lemma 4.1.1. *Let $(X_0, \Lambda_0) \in \mathbb{C}^{n \times m} \times \mathbb{C}^{m \times m}$ be a minimal invariant pair with minimality index at most ℓ and let $\mathbf{F}$ be defined as in* (4.4) *with W chosen such that $W^{\mathsf{H}} \mathbf{V}_\ell(X_0, \Lambda_0)$ is invertible. Then, for any nonsingular matrix $G \in \mathbb{C}^{m \times m}$, $W^{\mathsf{H}} \mathbf{V}_\ell(X_0 G, G^{-1} \Lambda_0 G)$ is invertible and*

$$\mathbf{F}_G(X, \Lambda) = \begin{bmatrix} \mathbf{T}(X, \Lambda) \\ W^{\mathsf{H}} \big[\mathbf{V}_\ell(X, \Lambda) - \mathbf{V}_\ell(X_0 G, G^{-1} \Lambda_0 G) \big] \end{bmatrix}$$

satisfies

$$\mathbf{F}_G(XG, G^{-1} \Lambda G) = \mathbf{F}(X, \Lambda) G \tag{4.5}$$

as well as

$$\mathrm{D}^k \mathbf{F}_G(XG, G^{-1} \Lambda G)(\triangle X G, G^{-1} \triangle \Lambda G)^k = \mathrm{D}^k \mathbf{F}(X, \Lambda)(\triangle X, \triangle \Lambda)^k G$$

for any pair $(X, \Lambda) \in C^{n \times m} \times \mathbb{C}^{m \times m}$ and any positive integer k.

Proof. The invertibility of $W^{\mathsf{H}} \mathbf{V}_\ell(X_0 G, G^{-1} \Lambda_0 G) = W^{\mathsf{H}} \mathbf{V}_\ell(X_0, \Lambda_0) G$ as well as the identity in (4.5) are direct consequences of Proposition 3.2.3. The statement about the derivatives then follows inductively from (4.5) by taking derivatives and applying the chain rule. □

4.1.1 Characterization of simple invariant pairs. In [83, Theorem 10], it has been shown that a minimal invariant pair (X_0, Λ_0) is simple if and only if $(X, \Lambda) = (X_0, \Lambda_0)$ is a regular solution of the nonlinear equation (4.3) in the sense that the derivative of $\mathbf{F}$ at (X_0, Λ_0) is invertible.

Theorem 4.1.2 ([83, Theorem 10])**.** *Let $(X_0, \Lambda_0) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ be a minimal invariant pair of some matrix-valued function T with minimality index at most ℓ and assume that $W^{\mathsf{H}}\mathbf{V}_\ell(X_0, \Lambda_0)$ is nonsingular. Then the Fréchet derivative $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ of $\mathbf{F}$ as defined in* (4.4) *at (X_0, Λ_0) is invertible if and only if (X_0, Λ_0) is simple.*

Theorem 4.1.2 has been proven in [83, Theorem 10]. Here, we will present a different proof based on extraction of minimal invariant pairs via Proposition 3.4.1 and a characterization of the null space of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ given in Theorem 4.1.3 below.

Proof of Theorem 4.1.2. Assume that (X_0, Λ_0) is simple and let $(\triangle X, \triangle\Lambda)$ be a null vector of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$. If $(\triangle X, \triangle\Lambda) \neq (0, 0)$, then Theorem 4.1.3 below yields the existence of a minimal invariant pair

$$(\hat{X},\, \hat{\Lambda}) = \left([X_0,\, Y],\, \begin{bmatrix} \Lambda_0 & Z \\ 0 & \Theta \end{bmatrix}\right)$$

of size larger than (X_0, Λ_0) such that $\operatorname{spec}\hat{\Lambda} = \operatorname{spec}\Lambda_0$, contradicting the fact that (X_0, Λ_0) is simple. Hence, $(\triangle X,\, \triangle\Lambda) = (0, 0)$, showing the invertibility of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$.

For the converse statement, let (X_0, Λ_0) be non-simple. Then, Proposition 3.4.3 yields vectors $\tilde{y}$, $\tilde{z}$, as well as an eigenvalue θ of Λ_0 such that $\bigl([X_0, \tilde{y}], \left[\begin{smallmatrix} \Lambda_0 & \tilde{z} \\ 0 & \theta \end{smallmatrix}\right]\bigr)$ is a minimal invariant pair of T. Let $\psi \neq 0$ be a left eigenvector of Λ_0 belonging to the eigenvalue θ. Then, by Theorem 4.1.3, there exist vectors y, z such that $\bigl([X_0, y], \left[\begin{smallmatrix} \Lambda_0 & z \\ 0 & \theta \end{smallmatrix}\right]\bigr)$ is a minimal invariant pair of T and $(\triangle X,\, \triangle\Lambda) = (y\psi^{\mathsf{H}},\, z\psi^{\mathsf{H}})$ is a null vector of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$. It remains to show that $(\triangle X, \triangle\Lambda) \neq (0, 0)$. Since $\psi \neq 0$, $(\triangle X,\, \triangle\Lambda) = (0, 0)$ would entail $(y, z) = (0, 0)$. But then the last column of $\mathbf{V}_k\bigl([X_0,\, y],\, \left[\begin{smallmatrix} \Lambda_0 & z \\ 0 & \theta \end{smallmatrix}\right]\bigr)$ would be zero for all $k \in \mathbb{N}$, contradicting the minimality of $\bigl([X_0, y], \left[\begin{smallmatrix} \Lambda_0 & z \\ 0 & \theta \end{smallmatrix}\right]\bigr)$. □

4.1.2 Characterization of non-simple invariant pairs. For a non-simple invariant pair (X_0, Λ_0), the Fréchet derivative $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ is not invertible by Theorem 4.1.2. Consequently, since $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ is a mapping between finite-dimensional vector spaces, the null space $\ker \mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ is non-trivial. In fact, the subsequent theorem reveals a close relationship between the null vectors of the Fréchet derivative $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ and extensions of the pair (X_0, Λ_0) having a smaller multiplicity.

Theorem 4.1.3. *Let $(X_0, \Lambda_0) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ be a minimal invariant pair of some matrix-valued function T with minimality index at most ℓ and assume that $W^{\mathsf{H}}\mathbf{V}_\ell(X_0, \Lambda_0)$ is invertible. Then every null vector $(\triangle X, \triangle\Lambda)$ of the Fréchet derivative $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ of $\mathbf{F}$ as defined in* (4.4) *at (X_0, Λ_0) has the form*

$$\bigl(\triangle X,\, \triangle\Lambda\bigr) = \bigl(Y\Psi^{\mathsf{H}},\, Z\Psi^{\mathsf{H}}\bigr), \qquad Y \in \mathbb{C}^{n\times r}, \quad Z, \Psi \in \mathbb{C}^{m\times r}, \quad r \geq 0, \tag{4.6}$$

where

(NV1) *Ψ has full column rank and $\Psi^{\mathsf{H}}\Lambda_0 = \Theta\Psi^{\mathsf{H}}$ for some suitable $\Theta \in \mathbb{C}^{r\times r}$,*

(NV2) $\left(\big[X_0,\, Y\big],\, \begin{bmatrix}\Lambda_0 & Z\\ 0 & \Theta\end{bmatrix}\right)$ *is a minimal invariant pair of T,*

(NV3) $W^{\mathsf{H}}\mathbf{V}_\ell\left(\big[X_0,\, Y\big],\, \begin{bmatrix}\Lambda_0 & Z\\ 0 & \Theta\end{bmatrix}\right) = \big[W^{\mathsf{H}}\mathbf{V}_\ell(X_0,\Lambda_0),\ 0\big].$

By convention, $r = 0$ corresponds to $(\triangle X, \triangle\Lambda) = (0, 0)$.

Conversely, any $(\triangle X, \triangle\Lambda)$ of the form (4.6) satisfying (NV1)–(NV3) is a null vector of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$. Moreover, if $\tilde Y, \tilde Z, \Psi$ satisfy (NV1)–(NV2), then there exists a matrix $\Phi \in \mathbb{C}^{m\times r}$ such that Y, Z, Ψ defined as

$$Y := \tilde Y - X_0\Phi, \qquad Z := \tilde Z - (\Lambda_0\Phi - \Phi\Theta) \tag{4.7}$$

satisfy (NV1)–(NV3).

Proof. Let $(\triangle X, \triangle\Lambda)$ be a null vector of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$, which is equivalent to the conditions

$$\begin{aligned}\mathrm{D}\mathbf{T}(X_0,\Lambda_0)(\triangle X, \triangle\Lambda) &= 0,\\ W^{\mathsf{H}}\mathrm{D}\mathbf{V}_\ell(X_0,\Lambda_0)(\triangle X, \triangle\Lambda) &= 0.\end{aligned} \tag{4.8}$$

By Proposition 3.3.9, the first condition implies that

$$(\hat X,\ \hat\Lambda) = \left(\big[X_0,\ \triangle X\big],\ \begin{bmatrix}\Lambda_0 & \triangle\Lambda\\ 0 & \Lambda_0\end{bmatrix}\right) \in \mathbb{C}^{n\times 2m} \times \mathbb{C}^{2m\times 2m}$$

constitutes an invariant pair of T. Now set $r := \operatorname{rank} \mathbf{V}_{2m}\big(\hat X, \hat\Lambda\big) - m$. Because $\mathbf{V}_{2m}(X_0, \Lambda_0)$ has full column rank, Proposition 3.3.7 implies that r is nonnegative. Furthermore, we find a matrix $B \in \mathbb{C}^{m\times m}$ and an invertible matrix $C \in \mathbb{C}^{m\times m}$ such that

$$\mathbf{V}_{2m}(\hat X,\ \hat\Lambda)\begin{bmatrix}I & BC\\ 0 & C\end{bmatrix} = \Big[\mathbf{V}_{2m}(X_0,\Lambda_0),\ VE_r^{\mathsf{H}}\Big],$$

where $V \in \mathbb{C}^{2mn\times r}$ and $E_r^{\mathsf{H}} = \big[I,\ 0\big] \in \mathbb{C}^{r\times m}$. Therefore, by Proposition 3.4.1,

$$\hat X\begin{bmatrix}I & BC\\ 0 & C\end{bmatrix} = \big[X_0,\ \tilde Y E_r^{\mathsf{H}}\big], \qquad \begin{bmatrix}I & BC\\ 0 & C\end{bmatrix}^{-1}\hat\Lambda\begin{bmatrix}I & BC\\ 0 & C\end{bmatrix} = \begin{bmatrix}\Lambda_0 & \tilde Z E_r^{\mathsf{H}}\\ 0 & \left[\begin{smallmatrix}\Theta & 0\\ * & *\end{smallmatrix}\right]\end{bmatrix}$$

with $\tilde Y \in \mathbb{C}^{n\times r}$, $\tilde Z \in \mathbb{C}^{m\times r}$, and $\Theta \in \mathbb{C}^{r\times r}$ such that $\big([X_0, \tilde Y],\, \left[\begin{smallmatrix}\Lambda_0 & \tilde Z\\ 0 & \Theta\end{smallmatrix}\right]\big)$ is a minimal invariant pair of T. The last block columns of the above equations amount to

$$X_0BC + \triangle XC = \tilde Y E_r^{\mathsf{H}}, \quad \Lambda_0BC + \triangle\Lambda C - B\Lambda_0C = \tilde Z E_r^{\mathsf{H}}, \quad C^{-1}\Lambda_0C = \left[\begin{smallmatrix}\Theta & 0\\ * & *\end{smallmatrix}\right].$$

This can be recast as

$$\triangle X = \tilde Y\Psi^{\mathsf{H}} - X_0B, \quad \triangle\Lambda = \tilde Z\Psi^{\mathsf{H}} - (\Lambda_0B - B\Lambda_0), \quad \Psi^{\mathsf{H}}\Lambda_0 = \Theta\Psi^{\mathsf{H}}, \tag{4.9}$$

where $\Psi := C^{-\mathsf{H}}E_r$ clearly has full column rank, thereby confirming (NV1). Using the special structure of $\triangle X$ and $\triangle\Lambda$ in (4.9), the second condition in (4.8) becomes

$$\begin{aligned}0 &= W^{\mathsf{H}}\mathrm{D}\mathbf{V}_\ell(X_0,\Lambda_0)\big(\tilde Y\Psi^{\mathsf{H}} - X_0B,\ \tilde Z\Psi^{\mathsf{H}} - (\Lambda_0B - B\Lambda_0)\big)\\ &= W^{\mathsf{H}}\big[\mathrm{D}\mathbf{V}_\ell(X_0,\Lambda_0)\big(\tilde Y\Psi^{\mathsf{H}},\ \tilde Z\Psi^{\mathsf{H}}\big) - \mathrm{D}\mathbf{V}_\ell(X_0,\Lambda_0)\big(X_0B,\ \Lambda_0B - B\Lambda_0\big)\big]\\ &= W^{\mathsf{H}}\big[\mathbf{V}_\ell(\tilde Y,\Theta) + \mathrm{D}_{[\Lambda_0,\Theta]}\mathbf{V}_\ell(X_0,\cdot)\tilde Z\big]\cdot\Psi^{\mathsf{H}} - W^{\mathsf{H}}\mathbf{V}_\ell(X_0,\Lambda_0)\cdot B.\end{aligned}$$

Solving the latter for B gives $B = \Phi\Psi^{\mathsf{H}}$, where

$$\Phi = \big[W^{\mathsf{H}}\mathbf{V}_\ell(X_0, \Lambda_0)\big]^{-1} W^{\mathsf{H}}\big[\mathbf{V}_\ell(\tilde{Y}, \Theta) + \mathrm{D}_{[\Lambda_0,\Theta]}\mathbf{V}_\ell(X_0, \cdot)\tilde{Z}\big], \tag{4.10}$$

showing that $(\triangle X, \triangle\Lambda)$ indeed has the form (4.6) with Y and Z defined as in (4.7). Moreover, by Proposition 3.2.3,

$$\left([X_0, Y], \begin{bmatrix}\Lambda_0 & Z\\ 0 & \Theta\end{bmatrix}\right) = \left([X_0, \tilde{Y}]\begin{bmatrix}I & -\Phi\\ 0 & I\end{bmatrix}, \begin{bmatrix}I & -\Phi\\ 0 & I\end{bmatrix}^{-1}\begin{bmatrix}\Lambda_0 & \tilde{Z}\\ 0 & \Theta\end{bmatrix}\begin{bmatrix}I & -\Phi\\ 0 & I\end{bmatrix}\right) \tag{4.11}$$

constitutes a minimal invariant pair of T, proving (NV2). Finally, with the aid of Proposition 3.3.7, (NV3) follows from

$$\begin{aligned}
&W^{\mathsf{H}}\mathbf{V}_\ell\left([X_0,\ Y],\ \begin{bmatrix}\Lambda_0 & Z\\ 0 & \Theta\end{bmatrix}\right) = W^{\mathsf{H}}\mathbf{V}_\ell\left([X_0,\ \tilde{Y}],\ \begin{bmatrix}\Lambda_0 & \tilde{Z}\\ 0 & \Theta\end{bmatrix}\right)\begin{bmatrix}I & -\Phi\\ 0 & I\end{bmatrix}\\
&= W^{\mathsf{H}}\Big[\mathbf{V}_\ell(X_0, \Lambda_0),\ \mathbf{V}_\ell(\tilde{Y}, \Theta) + \mathrm{D}_{[\Lambda_0,\Theta]}\mathbf{V}_\ell(X_0, \cdot)\tilde{Z}\Big]\begin{bmatrix}I & -\Phi\\ 0 & I\end{bmatrix}\\
&= W^{\mathsf{H}}\mathbf{V}_\ell(X_0, \Lambda_0)\,[I,\ \Phi]\begin{bmatrix}I & -\Phi\\ 0 & I\end{bmatrix} = \big[W^{\mathsf{H}}\mathbf{V}_\ell(X_0, \Lambda_0),\ 0\big].
\end{aligned} \tag{4.12}$$

Conversely, let $(\triangle X, \triangle\Lambda)$ be given by (4.6) and assume that (NV1)–(NV3) hold. Then, by Proposition 3.3.7, (NV2) and (NV3) imply

$$\begin{aligned}
\mathbf{T}(Y,\ \Theta) + \mathrm{D}_{[\Lambda_0,\Theta]}\mathbf{T}(X_0, \cdot)Z &= 0,\\
W^{\mathsf{H}}\big[\mathbf{V}_\ell(Y, \Theta) + \mathrm{D}_{[\Lambda_0,\Theta]}\mathbf{V}_\ell(X_0, \cdot)Z\big] &= 0.
\end{aligned} \tag{4.13}$$

Let $\mathcal{C}$ be a contour enclosing the eigenvalues of both Λ_0 and Θ in its interior. Since, by Lemma 3.3.8 and (NV1),

$$\begin{aligned}
\mathrm{D}\mathbf{T}(X_0, \Lambda_0)(\triangle X, \triangle\Lambda) &= \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}} T(\xi)\big[Y + X_0(\xi I - \Lambda_0)^{-1}Z\big]\Psi^{\mathsf{H}}(\xi I - \Lambda_0)^{-1}\,\mathrm{d}\xi\\
&= \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}} T(\xi)\big[Y + X_0(\xi I - \Lambda_0)^{-1}Z\big](\xi I - \Theta)^{-1}\,\mathrm{d}\xi \cdot \Psi^{\mathsf{H}}\\
&= \big[\mathbf{T}(Y, \Theta) + \mathrm{D}_{[\Lambda_0,\Theta]}\mathbf{T}(X_0, \cdot)Z\big]\cdot\Psi^{\mathsf{H}},
\end{aligned}$$

the first equation in (4.13) shows that the first condition in (4.8) holds. Likewise, the second condition in (4.8) follows from the second equation in (4.13) via an analogous calculation. Hence, $(\triangle X, \triangle\Lambda)$ is a null vector of $\mathrm{D}T(X_0, \Lambda_0)$ as claimed.

Lastly, let $\tilde{Y}, \tilde{Z}, \Psi$ be such that (NV1)–(NV2) are satisfied and define Y, Z via (4.7) with Φ as in (4.10). Then, clearly, Y, Z, Ψ satisfy (NV1). Additionally, (4.11) and (4.12) show that (NV2) and (NV3) are satisfied as well. □

4.1.3 Characterization of generic bifurcations. In the following, we confine ourselves to a special class of nonlinear eigenvalue problems, which we refer to as real.

Definition 4.1.4. A nonlinear eigenvalue problem, induced by a matrix-valued function $T : D \to \mathbb{C}^{n\times n}$, is called *real* if and only if $D \subset \mathbb{C}$ is closed under complex conjugation and $T(\overline{\lambda}) = \overline{T(\lambda)}$ for all $\lambda \in D$.

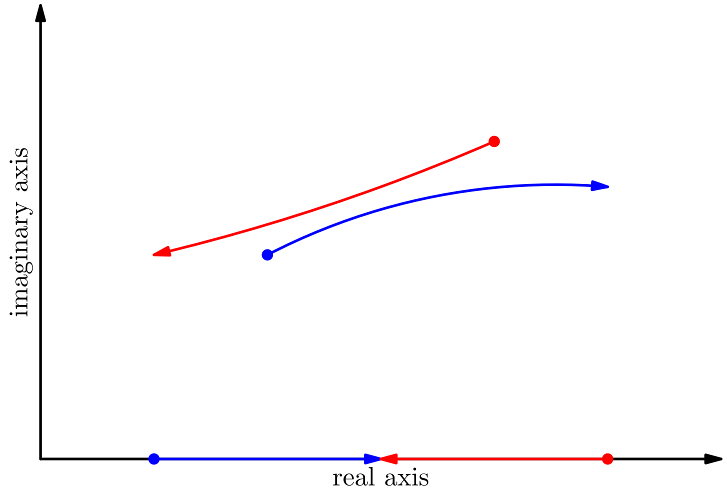

Figure 4.1: Illustration of the typical movement of eigenvalues under one-parameter variation: Eigenvalues on the real axis collide while eigenvalues in the complex plane miss each other.

For a linear eigenvalue problem, $T(\lambda) = \lambda I - A$, $D = \mathbb{C}$, Definition 4.1.4 coincides with the usual notion of realness, meaning that the matrix A has only real entries. It is well-known that the non-real eigenvalues of a real linear eigenvalue problem occur in complex conjugate pairs. This fact remains true for real nonlinear eigenvalue problems.

Theorem 4.1.5. *The spectrum of a real nonlinear eigenvalue problem is closed under complex conjugation.*

Proof. Let λ be an eigenvalue of the nonlinear eigenvalue problem and $x \neq 0$ a corresponding eigenvector. Then, $\overline{x} \neq 0$ and

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

showing that also $\overline{\lambda}$ is an eigenvalue. □

When dealing with real nonlinear eigenvalue problems, it is often reasonable to reflect their particular spectral structure in the minimal invariant pairs. That is, for every Jordan chain associated with an eigenvalue λ contained in a pair (X, Λ), this pair should also include a corresponding Jordan chain for the eigenvalue $\overline{\lambda}$. If the latter condition is met, the pair under consideration can be chosen real, i.e., $(X, \Lambda) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{m\times m}$, by a suitable similarity transform.

Intuitively, the most likely situation for a simple, real minimal invariant pair to become non-simple is when a real eigenvalue contained in the pair meets a real eigenvalue not contained in the pair; see Figure 4.1. This intuition has been made mathematically rigorous for linear eigenvalue problems already in the classic works by Arnol′d [5, 6].

From a more general perspective, it is well known that singular solutions in one-parameter systems occur at limit points (see [1, 20, 50]), where the tangent of the branch is vertical with respect to the parameter coordinate. In a generic sense, these limit points are *quadratic turning points*, which are defined by three nondegeneracy conditions. Applied to the nonlinear equation (4.3), the first two conditions read as follows.

(TP1) There is $(\triangle X_0, \triangle\Lambda_0) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{m\times m} \setminus \{(0,0)\}$ such that

$$\ker \mathrm{D}\mathbf{F}(X_0, \Lambda_0) = \operatorname{span}\bigl\{(\triangle X_0, \triangle\Lambda_0)\bigr\}.$$

(TP2) For $(\triangle X_0, \triangle\Lambda_0)$ as in (TP1),

$$\mathrm{D}^2\mathbf{F}(X_0, \Lambda_0)(\triangle X_0, \triangle\Lambda_0)^2 \notin \operatorname{im} \mathrm{D}\mathbf{F}(X_0, \Lambda_0).$$

The third condition, which will be discussed in Section 4.2.2, describes transversality with respect to the parameter.

Theorem 4.1.6. *Let $(X_0, \Lambda_0) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{m\times m}$ be a real minimal invariant pair with minimality index at most ℓ of a real nonlinear eigenvalue problem induced by the matrix-valued function T and let $\mathbf{F}$ be defined as in* (4.4) *with $W^{\mathsf{H}}\mathbf{V}_\ell(X_0, \Lambda_0)$ invertible. Then, the turning point conditions (TP1) and (TP2) above are equivalent to the following set of conditions.*

(J1) *The pair (X_0, Λ_0) has multiplicity* 2.

(J2) *T has a real eigenvalue μ of geometric multiplicity* 1 *such that*

$$\operatorname{alg}_T \mu = 2, \qquad \operatorname{alg}_{\Lambda_0} \mu = 1.$$

The original proof of Theorem 4.1.6 given in [21] proceeded in several steps. First, the statement of the theorem was shown for generalized linear eigenvalue problems, $T(\lambda) = \lambda B - A$, in [21, Lemma A.1] using the Kronecker canonical form of (A, B) and known results [81] concerning null spaces of generalized Sylvester operators. For the special case that the matrix B is invertible, a proof has already appeared in [22, Theorem 2.3]. Via linearization, the result is then extended to polynomial eigenvalue problems in [21, Lemma A.2]. Finally, the result is transferred to the general nonlinear setting through interpolation in [21, Theorem 3.1]. Here, we present a more direct proof for Theorem 4.1.6, which avoids the detour via polynomial and linear eigenvalue problems. It is based on the characterization of the null space of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ in Theorem 4.1.3 as well as the theory of minimal invariant pairs introduced in Chapter 3. Moreover, we will require the upcoming lemma.

Lemma 4.1.7. *Let $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$, $p \in \mathbb{C}^n$, $q \in \mathbb{C}^m$ and let $\psi \in \mathbb{C}^m$ be a left eigenvector of Λ corresponding to the eigenvalue μ. Then, for any $k \in \mathbb{N}$,*

$$\mathrm{D}^k\mathbf{T}(X, \Lambda)\bigl(p\psi^{\mathsf{H}}, q\psi^{\mathsf{H}}\bigr)^k = k \cdot \bigl(\psi^{\mathsf{H}} q\bigr)^{k-1} \cdot \frac{\partial^{k-1}}{\partial\theta^{k-1}} \bigl[T(\theta)p + \mathrm{D}_{[\Lambda,\theta]}\mathbf{T}(X, \cdot)q\bigr]\bigg|_{\theta=\mu} \cdot \psi^{\mathsf{H}}.$$

Proof. Inductively, one easily shows that

$$\frac{\partial^{k-1}}{\partial\theta^{k-1}} (\xi - \theta)^{-1}\bigg|_{\theta=\mu} = (k-1)! \cdot (\xi - \mu)^{-k}.$$

Therefore, and since $\psi^{\mathsf{H}}(\xi I - \Lambda)^{-1} = (\xi - \mu)^{-1}\psi^{\mathsf{H}}$, Lemma 3.3.8 implies

$$\mathrm{D}^k\mathbf{T}(X, \Lambda)(p\psi^{\mathsf{H}}, q\psi^{\mathsf{H}})^k = k \cdot \mathrm{D}_\Lambda^{k-1}\mathbf{T}(p\psi^{\mathsf{H}}, \Lambda)(q\psi^{\mathsf{H}})^{k-1} + \mathrm{D}_\Lambda^k\mathbf{T}(X, \Lambda)(q\psi^{\mathsf{H}})^k,$$

where

$$\begin{aligned}
\mathrm{D}_\Lambda^{k-1}\mathbf{T}(p\psi^{\mathsf{H}},\Lambda)(q\psi^{\mathsf{H}})^{k-1} &= \frac{(k-1)!}{2\pi\mathrm{i}}\int_\Gamma T(\xi)p\psi^{\mathsf{H}}(\xi I-\Lambda)^{-1}\bigl[q\psi^{\mathsf{H}}(\xi I-\Lambda)^{-1}\bigr]^{k-1}\,\mathrm{d}\xi \\
&= \frac{(k-1)!}{2\pi\mathrm{i}}\int_\Gamma T(\xi)p(\xi-\mu)^{-1}\psi^{\mathsf{H}}\bigl[q(\xi-\mu)^{-1}\psi^{\mathsf{H}}\bigr]^{k-1}\,\mathrm{d}\xi \\
&= (\psi^{\mathsf{H}}q)^{k-1}\cdot\frac{(k-1)!}{2\pi\mathrm{i}}\int_\Gamma T(\xi)p(\xi-\mu)^{-k}\,\mathrm{d}\xi\cdot\psi^{\mathsf{H}} \\
&= (\psi^{\mathsf{H}}q)^{k-1}\cdot\frac{\partial^{k-1}}{\partial\theta^{k-1}}T(\theta)p\bigg|_{\theta=\mu}\cdot\psi^{\mathsf{H}}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{D}_\Lambda^{k}\mathbf{T}(X,\Lambda)(q\psi^{\mathsf{H}})^{k} &= \frac{k!}{2\pi\mathrm{i}}\int_\Gamma T(\xi)X(\xi I-\Lambda)^{-1}\bigl[q\psi^{\mathsf{H}}(\xi I-\Lambda)^{-1}\bigr]^{k}\,\mathrm{d}\xi \\
&= \frac{k!}{2\pi\mathrm{i}}\int_\Gamma T(\xi)X(\xi I-\Lambda)^{-1}\bigl[q(\xi-\mu)^{-1}\psi^{\mathsf{H}}\bigr]^{k}\,\mathrm{d}\xi \\
&= (\psi^{\mathsf{H}}q)^{k-1}\cdot\frac{k!}{2\pi\mathrm{i}}\int_\Gamma T(\xi)X(\xi I-\Lambda)^{-1}q(\xi-\mu)^{-k}\,\mathrm{d}\xi\cdot\psi^{\mathsf{H}} \\
&= k\cdot(\psi^{\mathsf{H}}q)^{k-1}\cdot\frac{\partial^{k-1}}{\partial\theta^{k-1}}\mathrm{D}_{[\Lambda,\theta]}\mathbf{T}(X,\cdot)q\bigg|_{\theta=\mu}\cdot\psi^{\mathsf{H}}
\end{aligned}$$

with a contour Γ enclosing the eigenvalues of Λ in its interior. Putting everything together, the assertion follows. □

Proof of Theorem 4.1.6. Due to Lemma 4.1.1 we may, w.l.o.g., assume throughout the entire proof that Λ_0 is in Jordan canonical form. The statement for arbitrary Λ_0 then follows by an adequate transformation.

To begin with, assume that (J1)–(J2) hold. By (J2), Λ_0 has a simple, real eigenvalue μ. W.l.o.g., we may assume the corresponding Jordan block to be located in the top left corner; i.e.,

$$\Lambda_0 = \begin{bmatrix}\mu & 0\\ 0 & \Lambda_2\end{bmatrix}. \tag{4.14}$$

Moreover, since the multiplicity of (X_0,Λ_0) is two and T possesses a Jordan chain of length two associated with μ, there is a vector $\tilde{y}$ such that $\bigl([X_0,\tilde{y}],\left[\begin{smallmatrix}\Lambda_0 & e_1\\ 0 & \mu\end{smallmatrix}\right]\bigr)$ is a simple invariant pair of T. Taking the obvious left eigenvector e_1 of Λ_0 associated with the eigenvalue μ, Theorem 4.1.3 yields the existence of $y=\tilde{y}-X_0\phi$ and $z=e_1-(\Lambda_0-\mu I)\phi$ such that

(P1) $\bigl(ye_1^{\mathsf{H}},ze_1^{\mathsf{H}}\bigr)\in\ker\mathrm{D}\mathbf{F}(X_0,\Lambda_0)$;

(P2) $\bigl([X_0,\,y],\,\left[\begin{smallmatrix}\Lambda_0 & z\\ 0 & \mu\end{smallmatrix}\right]\bigr)$ is a minimal invariant pair of T, which—by comparing multiplicities—is again simple;

(P3) $\bigl[W^{\mathsf{H}}\mathbf{V}_\ell(X_0,\Lambda_0)\bigr]^{-1}W^{\mathsf{H}}\mathbf{V}_\ell\bigl([X_0,\,y],\,\left[\begin{smallmatrix}\Lambda_0 & z\\ 0 & \mu\end{smallmatrix}\right]\bigr)=\bigl[I,\,0\bigr]$.

We will now show that $\ker\mathrm{D}\mathbf{F}(X_0,\Lambda_0)=\operatorname{span}\{(ye_1^{\mathsf{H}},ze_1^{\mathsf{H}})\}$. Since $\ker\mathrm{D}\mathbf{F}(X_0,\Lambda_0)$ is a vector space, property (P1) implies $\operatorname{span}\{(ye_1^{\mathsf{H}},ze_1^{\mathsf{H}})\}\subset\ker\mathrm{D}\mathbf{F}(X_0,\Lambda_0)$. For the opposite inclusion, let $(\triangle X,\triangle\Lambda)\in\ker\mathrm{D}\mathbf{F}(X_0,\Lambda_0)$ be arbitrary. Then, according to Theorem 4.1.3, $(\triangle X,\triangle\Lambda)=(Y\Psi^{\mathsf{H}},Z\Psi^{\mathsf{H}})$, where $Y\in\mathbb{C}^{n\times r}$, $Z,\Psi\in\mathbb{C}^{m\times r}$, and

(P4) Ψ has full column rank, $\Psi^{\mathsf{H}}\Lambda_0 = \Theta\Psi^{\mathsf{H}}$;

(P5) $\big([X_0,\, Y],\, \left[\begin{smallmatrix}\Lambda_0 & Z\\ 0 & \Theta\end{smallmatrix}\right]\big)$ is a minimal invariant pair of T;

(P6) $\big[W^{\mathsf{H}}\mathbf{V}_\ell(X_0,\Lambda_0)\big]^{-1}W^{\mathsf{H}}\mathbf{V}_\ell\big([X_0,\, Y],\, \left[\begin{smallmatrix}\Lambda_0 & Z\\ 0 & \Theta\end{smallmatrix}\right]\big) = \big[I,\, 0\big]$.

Because (X_0, Λ_0) is of multiplicity two, property (P5) entails $r \leq 1$. If $r = 0$, then $(\triangle X, \triangle\Lambda) = (0, 0) \in \operatorname{span}\{(ye_1^{\mathsf{H}}, ze_1^{\mathsf{H}})\}$. Hence, we are left with the case $r = 1$. In this case, it is apparent from property (P5) that $\Theta = \mu$. Consequently, by (P4), Ψ is a left eigenvector of Λ_0 in (4.14) associated with the eigenvalue μ, and hence, $\Psi = \alpha e_1$ for some suitable $\alpha \in \mathbb{C}$. Furthermore, Proposition 3.4.2 yields the existence of a matrix $\left[\begin{smallmatrix}S_{11} & S_{12}\\ S_{21} & s_{22}\end{smallmatrix}\right] \in \mathbb{C}^{(m+1)\times(m+1)}$ such that

$$\big[X_0,\, Y\big] = \big[X_0,\, y\big]\begin{bmatrix}S_{11} & S_{12}\\ S_{21} & s_{22}\end{bmatrix}, \quad \begin{bmatrix}S_{11} & S_{12}\\ S_{21} & s_{22}\end{bmatrix}\begin{bmatrix}\Lambda_0 & Z\\ 0 & \Theta\end{bmatrix} = \begin{bmatrix}\Lambda_0 & z\\ 0 & \mu\end{bmatrix}\begin{bmatrix}S_{11} & S_{12}\\ S_{21} & s_{22}\end{bmatrix}.$$

Therefore, by properties (P3) and (P6),

$$\begin{aligned}\big[I,\, 0\big] &= \big[W^{\mathsf{H}}\mathbf{V}_\ell(X_0,\Lambda_0)\big]^{-1}W^{\mathsf{H}}\mathbf{V}_\ell\big([X_0,\, Y],\, \left[\begin{smallmatrix}\Lambda_0 & Z\\ 0 & \Theta\end{smallmatrix}\right]\big)\\ &= \big[W^{\mathsf{H}}\mathbf{V}_\ell(X_0,\Lambda_0)\big]^{-1}W^{\mathsf{H}}\mathbf{V}_\ell\big([X_0,\, y],\, \left[\begin{smallmatrix}\Lambda_0 & z\\ 0 & \mu\end{smallmatrix}\right]\big)\left[\begin{smallmatrix}S_{11} & S_{12}\\ S_{21} & s_{22}\end{smallmatrix}\right]\\ &= \big[I,\, 0\big]\left[\begin{smallmatrix}S_{11} & S_{12}\\ S_{21} & s_{22}\end{smallmatrix}\right] = \big[S_{11},\, S_{12}\big].\end{aligned}$$

Combining the last two identities yields $Y = s_{22}\cdot y$ and $Z = s_{22}\cdot z$. Consequently, $(\triangle X, \triangle\Lambda) = (\alpha s_{22})\cdot(ye_1^{\mathsf{H}}, ze_1^{\mathsf{H}}) \in \operatorname{span}\{(ye_1^{\mathsf{H}}, ze_1^{\mathsf{H}})\}$, proving (TP1).

The proof of (TP2) is by contradiction. To this end, assume that there exists $(\triangle X, \triangle\Lambda) \in \mathbb{C}^{n\times m}\times\mathbb{C}^{m\times m}$ such that

$$\mathrm{D}^2\mathbf{F}(X_0,\Lambda_0)\big(ye_1^{\mathsf{H}}, ze_1^{\mathsf{H}}\big)^2 = \mathrm{D}\mathbf{F}(X_0,\Lambda_0)(\triangle X, \triangle\Lambda).$$

Then, Proposition 3.3.9 implies that

$$(\hat{X}, \hat{\Lambda}) = \left(\Big[X_0,\, ye_1^{\mathsf{H}},\, -\tfrac{1}{2}\triangle X\Big],\, \begin{bmatrix}\Lambda_0 & ze_1^{\mathsf{H}} & -\frac{1}{2}\triangle\Lambda\\ 0 & \Lambda_0 & ze_1^{\mathsf{H}}\\ 0 & 0 & \Lambda_0\end{bmatrix}\right)$$

constitutes an invariant pair of T. We will now apply several basis transformations to this invariant pair. The first transformation uses the matrix $S_1 = \left[\begin{smallmatrix}I & \phi e_1^{\mathsf{H}} & \phi e_1^{\mathsf{H}}\phi e_1^{\mathsf{H}}\\ 0 & I & \phi e_1^{\mathsf{H}}\\ 0 & 0 & I\end{smallmatrix}\right]$ and leads to the equivalent invariant pair

$$\big(\hat{X}S_1,\, S_1^{-1}\hat{\Lambda}S_1\big) = \left(\Big[X_0,\, \tilde{y}e_1^{\mathsf{H}},\, -\tfrac{1}{2}\triangle X + \tilde{y}e_1^{\mathsf{H}}\phi e_1^{\mathsf{H}}\Big],\, \left[\begin{smallmatrix}\Lambda_0 & e_1e_1^{\mathsf{H}} & -\frac{1}{2}\triangle\Lambda - (I - e_1e_1^{\mathsf{H}})\phi e_1^{\mathsf{H}}\\ 0 & \Lambda_0 & e_1e_1^{\mathsf{H}}\\ 0 & 0 & \Lambda_0\end{smallmatrix}\right]\right)$$

if we use that $e_1^{\mathsf{H}}\Lambda_0 = \mu e_1^{\mathsf{H}}$, $y + X_0\phi = \tilde{y}$, and $z + \Lambda_0\phi - \mu\phi = e_1$. Taking into account the block structure of Λ_0 in (4.14), the pair $\big(\hat{X}S_1,\, S_1^{-1}\hat{\Lambda}S_1\big)$ can also be written in the form

$$\left(\Big[x_1,\, X_2,\, \tilde{y},\, 0,\, -\tfrac{1}{2}h_1 + \tilde{y}\phi_1,\, -\tfrac{1}{2}H_2\Big],\, \left[\begin{smallmatrix}\mu & 0 & 1 & 0 & -\frac{1}{2}\delta_{11} & -\frac{1}{2}\Delta_{12}\\ 0 & \Lambda_2 & 0 & 0 & -\frac{1}{2}\Delta_{21} - \phi_2 & -\frac{1}{2}\Delta_{22}\\ 0 & 0 & \mu & 0 & 1 & 0\\ 0 & 0 & 0 & \Lambda_2 & 0 & 0\\ 0 & 0 & 0 & 0 & \mu & 0\\ 0 & 0 & 0 & 0 & 0 & \Lambda_2\end{smallmatrix}\right]\right),$$

where we have partitioned $X_0 = [x_1, X_2]$, $\triangle X = [h_1, H_2]$, $\triangle\Lambda = \left[\begin{smallmatrix} \delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{smallmatrix}\right]$, and $\phi = \left[\begin{smallmatrix} \phi_1 \\ \phi_2 \end{smallmatrix}\right]$ conformally. Performing a perfect shuffle, we obtain the invariant pair

$$\left(\Big[x_1,\ \tilde{y},\ -\tfrac{1}{2}h_1 + \tilde{y}\phi_1,\ X_2,\ 0,\ -\tfrac{1}{2}H_2\Big],\ \begin{bmatrix} \mu & 1 & -\frac{1}{2}\delta_{11} & 0 & 0 & -\frac{1}{2}\Delta_{12} \\ 0 & \mu & 1 & 0 & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2}\Delta_{21} - \phi_2 & \Lambda_2 & 0 & -\frac{1}{2}\Delta_{22} \\ 0 & 0 & 0 & 0 & \Lambda_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Lambda_2 \end{bmatrix}\right).$$

Note that x_1 is an eigenvector of T belonging to the eigenvalue μ and therefore $x_1 \neq 0$. Since μ is not an eigenvalue of Λ_2, there exists a unique vector τ satisfying $(\Lambda_2 - \mu I)\tau = \frac{1}{2}\Delta_{21} + \phi_2$. Using this vector, we can bring the invariant pair into the upper block triangular form

$$\left(\Big[x_1,\ \tilde{y},\ -\tfrac{1}{2}h_1 + \tilde{y}\phi_1 + X_2\tau,\ X_2,\ 0,\ -\tfrac{1}{2}H_2\Big],\ \begin{bmatrix} \mu & 1 & -\frac{1}{2}\delta_{11} & 0 & 0 & -\frac{1}{2}\Delta_{12} \\ 0 & \mu & 1 & 0 & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \Lambda_2 & 0 & -\frac{1}{2}\Delta_{22} \\ 0 & 0 & 0 & 0 & \Lambda_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Lambda_2 \end{bmatrix}\right)$$

by transforming it using the matrix $S_2 = \operatorname{diag}\big(\left[\begin{smallmatrix} 1 & 0 \\ 0 & I \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 & 0 \\ \tau & I \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 & 0 \\ 0 & I \end{smallmatrix}\right]\big)$. With the aid of Proposition (3.3.7), we can extract the smaller invariant pair

$$(X_*,\ \Lambda_*) = \left(\Big[x_1,\ \tilde{y},\ -\tfrac{1}{2}h_1 + \tilde{y}\phi_1 + X_2\tau\Big],\ \begin{bmatrix} \mu & 1 & -\frac{1}{2}\delta_{11} \\ 0 & \mu & 1 \\ 0 & 0 & \mu \end{bmatrix}\right)$$

from the leading part of the upper block triangular form. A final transformation using the matrix $S_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{1}{2}\delta_{11} \\ 0 & 0 & 1 \end{bmatrix}$ then shows that

$$(X_* S_3,\ S_3^{-1}\Lambda_* S_3) = \left(\Big[x_1,\ \tilde{y},\ -\tfrac{1}{2}h_1 + \tilde{y}(\phi_1 + \tfrac{1}{2}\delta_{11}) + X_2\tau\Big],\ \begin{bmatrix} \mu & 1 & 0 \\ 0 & \mu & 1 \\ 0 & 0 & \mu \end{bmatrix}\right)$$

constitutes an invariant pair of T. By virtue of Lemma 3.1.11, the invariance of the last pair implies that T has a Jordan chain of length at least three associated with μ, in contradiction to the assumption that the algebraic multiplicity of μ as an eigenvalue of T is only two. Thus, (TP2) must hold.

To prove the converse statement, assume that (TP1) and (TP2) are satisfied. Because the null space of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ is nontrivial, the pair (X_0, Λ_0) cannot be simple by Theorem 4.1.2. Consequently, Proposition 3.4.3 implies the existence of vectors y, z and an eigenvalue μ of Λ_0 such that $\big([X_0,\, y],\, \left[\begin{smallmatrix} \Lambda_0 & z \\ 0 & \mu \end{smallmatrix}\right]\big)$ is a minimal invariant pair of T. Let $\psi \neq 0$ be a left eigenvector of Λ_0 belonging to μ and assume, w.l.o.g., that condition (NV3) of Theorem 4.1.3 is already satisfied so that $(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}}) \in \ker \mathrm{D}\mathbf{F}(X_0, \Lambda_0)$.

To establish (J1), we need to show that $\big([X_0,\, y],\, \left[\begin{smallmatrix} \Lambda_0 & z \\ 0 & \mu \end{smallmatrix}\right]\big)$ is simple. The proof is again by contradiction, so assume that there are vectors $\hat{y}$, $\hat{z}$, an eigenvalue ν of Λ_0, and a scalar α such that $\Big([X_0,\, y,\, \hat{y}],\, \left[\begin{smallmatrix} \Lambda_0 & z & \hat{z} \\ 0 & \mu & \alpha \\ 0 & 0 & \nu \end{smallmatrix}\right]\Big)$ is a minimal invariant pair of T. Moreover, assume, w.l.o.g., that the vectors are chosen such that

$$W^{\mathsf{H}}\mathbf{V}_\ell\left([X_0,\, y,\, \hat{y}],\, \left[\begin{smallmatrix} \Lambda_0 & z & \hat{z} \\ 0 & \mu & \alpha \\ 0 & 0 & \nu \end{smallmatrix}\right]\right) = \big[W^{\mathsf{H}}\mathbf{V}_\ell(X_0, \Lambda_0),\, 0,\, 0\big]. \tag{4.15}$$

If $\nu \neq \mu$, we may additionally choose $\alpha = 0$. Let $\zeta \neq 0$ be a left eigenvector of Λ_0 belonging to the eigenvalue ν. In case $\alpha = 0$, we will show that $(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}})$

and $(\hat{y}\zeta^{\mathsf{H}}, \hat{z}\zeta^{\mathsf{H}})$ are two linearly independent null vectors of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$, contradicting (TP1). To this end, let $\beta_1, \beta_2 \in \mathbb{C}$ be such that $\beta_1(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}}) + \beta_2(\hat{y}\zeta^{\mathsf{H}}, \hat{z}\zeta^{\mathsf{H}}) = 0$. Then, for all $k \in \mathbb{N}$,

$$[X_0,\ y,\ \hat{y}] \begin{bmatrix} \Lambda_0 & z & \hat{z} \\ 0 & \mu & 0 \\ 0 & 0 & \nu \end{bmatrix}^k \begin{bmatrix} 0 \\ \beta_1\psi^{\mathsf{H}} \\ \beta_2\zeta^{\mathsf{H}} \end{bmatrix} = [X_0,\ y,\ \hat{y}] \begin{bmatrix} 0 \\ \beta_1\psi^{\mathsf{H}} \\ \beta_2\zeta^{\mathsf{H}} \end{bmatrix} \Lambda_0^k = 0,$$

which combines to $\mathbf{V}_{m+2}\Big([X_0,\ y,\ \hat{y}],\ \left[\begin{smallmatrix} \Lambda_0 & z & \hat{z} \\ 0 & \mu & 0 \\ 0 & 0 & \nu \end{smallmatrix}\right]\Big) \left[\begin{smallmatrix} 0 \\ \beta_1\psi^{\mathsf{H}} \\ \beta_2\zeta^{\mathsf{H}} \end{smallmatrix}\right] = 0$. Due to the minimality of the pair $\Big([X_0,\ y,\ \hat{y}],\ \left[\begin{smallmatrix} \Lambda_0 & z & \hat{z} \\ 0 & \mu & 0 \\ 0 & 0 & \nu \end{smallmatrix}\right]\Big)$, the latter implies $\beta_1\psi^{\mathsf{H}} = \beta_2\zeta^{\mathsf{H}} = 0$, leading to $\beta_1 = \beta_2 = 0$ because $\psi, \zeta \neq 0$. To see that $(\hat{y}\zeta^{\mathsf{H}}, \hat{z}\zeta^{\mathsf{H}}) \in \mathrm{D}\mathbf{F}(X_0, \Lambda_0)$, extract the minimal invariant pair $\big([X_0,\ \hat{y}],\ \left[\begin{smallmatrix} \Lambda_0 & \hat{z} \\ 0 & \nu \end{smallmatrix}\right]\big)$ from $\Big([X_0,\ y,\ \hat{y}],\ \left[\begin{smallmatrix} \Lambda_0 & z & \hat{z} \\ 0 & \mu & 0 \\ 0 & 0 & \nu \end{smallmatrix}\right]\Big)$ and apply Theorem 4.1.3.

In the remaining case $\nu = \mu$, $\alpha \neq 0$, we can always achieve $\alpha = 1$ via a suitable transformation without sacrificing the normalization condition (4.15). By Proposition 3.3.7, the last column of $\mathbf{T}\Big([X_0,\ y,\ \hat{y}],\ \left[\begin{smallmatrix} \Lambda_0 & z & \hat{z} \\ 0 & \mu & 1 \\ 0 & 0 & \mu \end{smallmatrix}\right]\Big) = 0$ then amounts to

$$0 = T(\mu)\hat{y} + \frac{1}{2\pi\mathrm{i}} \int_\Gamma T(\xi)[X_0,\ y] \begin{bmatrix} \xi I - \Lambda_0 & -z \\ 0 & \xi - \mu \end{bmatrix}^{-1} \begin{bmatrix} \hat{z} \\ 1 \end{bmatrix} (\xi - \mu)^{-1}\,\mathrm{d}\xi$$

or, by inverting the block matrix,

$$\begin{aligned} 0 &= T(\mu)\hat{y} + \int_\Gamma \frac{T(\xi)}{2\pi\mathrm{i}} \Big[X_0(\xi I - \Lambda_0)^{-1}\big(\hat{z}(\xi-\mu)^{-1} + z(\xi-\mu)^{-2}\big) + y(\xi-\mu)^{-2}\Big]\,\mathrm{d}\xi \\ &= T(\mu)\hat{y} + \mathrm{D}_{[\Lambda_0,\mu]}\mathbf{T}(X_0,\cdot)\hat{z} + \frac{\partial}{\partial\theta}\mathrm{D}_{[\Lambda_0,\theta]}\mathbf{T}(X_0,\cdot)z\Big|_{\theta=\mu} + \dot{T}(\mu)y, \end{aligned}$$

where Γ is a contour enclosing the eigenvalues of Λ_0 in its interior. Furthermore, Lemma 4.1.7 implies

$$\mathrm{D}^2\mathbf{T}(X_0, \Lambda_0)(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}})^2 = 2\psi^{\mathsf{H}}z \cdot \Big[\dot{T}(\mu)y + \frac{\partial}{\partial\theta}\mathrm{D}_{[\Lambda_0,\theta]}\mathbf{T}(X_0,\cdot)z\Big|_{\theta=\mu}\Big] \cdot \psi^{\mathsf{H}}$$

and

$$\mathrm{D}\mathbf{T}(X_0, \Lambda_0)(\hat{y}\psi^{\mathsf{H}}, \hat{z}\psi^{\mathsf{H}}) = \Big[T(\mu)\hat{y} + \mathrm{D}_{[\Lambda_0,\mu]}\mathbf{T}(X_0,\cdot)\hat{z}\Big] \cdot \psi^{\mathsf{H}}$$

Combining the last three identities shows that $(\triangle X, \triangle\Lambda) = -2\psi^{\mathsf{H}}z \cdot (\hat{y}\psi^{\mathsf{H}}, \hat{z}\psi^{\mathsf{H}})$ satisfies $\mathrm{D}^2\mathbf{T}(X_0, \Lambda_0)(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}})^2 = \mathrm{D}\mathbf{T}(X_0, \Lambda_0)(\triangle X, \triangle\Lambda)$. A similar calculation using the last column of (4.15) demonstrates that $W^{\mathsf{H}}\mathrm{D}^2\mathbf{V}_\ell(X_0, \Lambda_0)(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}})^2 = W^{\mathsf{H}}\mathrm{D}\mathbf{V}_\ell(X_0, \Lambda_0)(\triangle X, \triangle\Lambda)$, and combining the two results yields

$$\mathrm{D}^2\mathbf{F}(X_0, \Lambda_0)(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}})^2 = \mathrm{D}\mathbf{F}(X_0, \Lambda_0)(\triangle X, \triangle\Lambda), \tag{4.16}$$

in contradiction to (TP2). Since both possible cases lead to contradictions, we have shown that $\big([X_0,\ y],\ \left[\begin{smallmatrix} \Lambda_0 & z \\ 0 & \mu \end{smallmatrix}\right]\big)$ constitutes a simple invariant pair of T, thereby establishing (J1).

We now turn to (J2). First of all, we observe that μ is bound to be real. Otherwise, $\Lambda_0 \in \mathbb{R}^{m \times m}$ and—because of the realness assumption on T—also the simple invariant pair $\big([X_0,\ y],\ \left[\begin{smallmatrix} \Lambda_0 & z \\ 0 & \mu \end{smallmatrix}\right]\big)$ must contain identical Jordan blocks for μ and $\overline{\mu}$.

This, however, cannot be true since the algebraic multiplicity of μ is raised by one in the transition from Λ_0 to $\left[\begin{smallmatrix}\Lambda_0 & z\\ 0 & \mu\end{smallmatrix}\right]$, whereas the algebraic multiplicity of $\overline{\mu}$ remains constant.

If μ in $\left[\begin{smallmatrix}\Lambda_0 & z\\ 0 & \mu\end{smallmatrix}\right]$ is uncoupled from the Jordan blocks in Λ_0, i.e., if $z = (\Lambda_0 - \mu I)\phi$ for some $\phi \in \mathbb{C}^m$, then $\psi^{\mathsf{H}} z = \psi^{\mathsf{H}}(\Lambda_0 - \mu I)\phi = 0$. Consequently, by Lemma 4.1.7, $\mathrm{D}^2\mathbf{T}(X_0, \Lambda_0)(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}})^2 = 0$. Additionally, via an analogous calculation, also $W^{\mathsf{H}}\mathrm{D}^2\mathbf{V}_\ell(X_0, \Lambda_0)(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}})^2 = 0$, showing that $(\triangle X, \triangle\Lambda) = (0, 0)$ is a solution of Equation (4.16), in contradiction to (TP2). Hence, μ is coupled to a Jordan block in Λ_0.

W.l.o.g., assume that the Jordan block to which μ is coupled is located in the lower, right corner of Λ_0. The associated left eigenvector is $\psi = e_m$. If $\hat{\psi} \neq 0$ is a left eigenvector of Λ_0 corresponding to a different Jordan block associated with μ, then $(y\psi^{\mathsf{H}}, z\psi^{\mathsf{H}})$ and $(y\hat{\psi}^{\mathsf{H}}, z\hat{\psi}^{\mathsf{H}})$ are two linearly independent null vectors of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$, contradicting (TP1). Hence, Λ_0 possesses only one Jordan block belonging to μ.

If the size of the Jordan block corresponding to μ is at least 2×2, then $e_{m-1}^{\mathsf{H}}\Lambda_0 = \mu e_{m-1}^{\mathsf{H}} + e_m^{\mathsf{H}}$, which can be rearranged into

$$\begin{aligned} e_{m-1}^{\mathsf{H}}(\xi I - \Lambda_0)^{-1} &= (\xi - \mu)^{-1} e_{m-1}^{\mathsf{H}} + (\xi - \mu)^{-1} e_m^{\mathsf{H}}(\xi I - \Lambda_0)^{-1} \\ &= (\xi - \mu)^{-1} e_{m-1}^{\mathsf{H}} + (\xi - \mu)^{-2} e_m^{\mathsf{H}}. \end{aligned}$$

Using this identity together with Lemma 3.3.8, we find

$$\begin{aligned} &\mathrm{D}\mathbf{T}(X_0, \Lambda_0)\big(y e_{m-1}^{\mathsf{H}}, z e_{m-1}^{\mathsf{H}}\big) \\ &= \frac{1}{2\pi\mathrm{i}} \int_\Gamma T(\xi)\big[y + X_0(\xi I - \Lambda_0)^{-1} z\big] e_{m-1}^{\mathsf{H}}(\xi I - \Lambda_0)^{-1}\,\mathrm{d}\xi \\ &= \frac{1}{2\pi\mathrm{i}} \int_\Gamma T(\xi)\big[y + X_0(\xi I - \Lambda_0)^{-1} z\big]\big[(\xi - \mu)^{-1} e_{m-1}^{\mathsf{H}} + (\xi - \mu)^{-2} e_m^{\mathsf{H}}\big]\,\mathrm{d}\xi \\ &= \big[T(\mu)y + \mathrm{D}_{[\Lambda_0, \mu]}\mathbf{T}(X_0, \cdot)z\big] \cdot e_{m-1}^{\mathsf{H}} + \frac{\partial}{\partial\theta}\big[T(\theta)y + \mathrm{D}_{[\Lambda_0, \theta]}\mathbf{T}(X_0, \cdot)z\big]\bigg|_{\theta=\mu} \cdot e_m^{\mathsf{H}}. \end{aligned}$$

The contour Γ needs to be chosen such that it encloses the eigenvalues of Λ_0 in its interior. The first summand in the above expansion of $\mathrm{D}\mathbf{T}(X_0, \Lambda_0)(y e_{m-1}^{\mathsf{H}}, z e_{m-1}^{\mathsf{H}})$ vanishes by Proposition 3.3.7 thanks to the invariance of the pair $\big([X_0,\ y],\ \left[\begin{smallmatrix}\Lambda_0 & z\\ 0 & \mu\end{smallmatrix}\right]\big)$. Comparing the remaining summand with the expression obtained by Lemma 4.1.7 for $\mathrm{D}^2\mathbf{T}(X_0, \Lambda_0)(y e_m^{\mathsf{H}}, z e_m^{\mathsf{H}})^2$ shows that $(\triangle X, \triangle\Lambda) = 2 e_m^{\mathsf{H}} z \cdot (y e_{m-1}^{\mathsf{H}}, z e_{m-1}^{\mathsf{H}})$ solves $\mathrm{D}^2\mathbf{T}(X_0, \Lambda_0)(y e_m^{\mathsf{H}}, z e_m^{\mathsf{H}})^2 = \mathrm{D}\mathbf{T}(X_0, \Lambda_0)(\triangle X, \triangle\Lambda)$. By a similar argumentation, also $W^{\mathsf{H}}\mathrm{D}^2\mathbf{V}_\ell(X_0, \Lambda_0)(y e_m^{\mathsf{H}}, z e_m^{\mathsf{H}})^2 = W^{\mathsf{H}}\mathrm{D}\mathbf{V}_\ell(X_0, \Lambda_0)(\triangle X, \triangle\Lambda)$ so that, in total, $(\triangle X, \triangle\Lambda)$ is a solution of Equation (4.16), contradicting (TP2). Consequently, the Jordan block of Λ_0 belonging to μ is of size 1×1, confirming that μ is a simple eigenvalue of Λ_0 and, as an eigenvalue of $\left[\begin{smallmatrix}\Lambda_0 & z\\ 0 & \mu\end{smallmatrix}\right]$, has algebraic multiplicity 2 and geometric multiplicity 1. The proof is completed by noting that the multiplicities of μ as an eigenvalue of T are identical to the latter because $\big([X_0,\ y],\ \left[\begin{smallmatrix}\Lambda_0 & z\\ 0 & \mu\end{smallmatrix}\right]\big)$ is a simple invariant pair. □

The conditions (J1) and (J2) in Theorem 4.1.6 state that there exists a Jordan chain x_0, x_1 of length two belonging to a real eigenvalue μ whose first vector x_0 (the eigenvector) is represented in the invariant pair (X_0, Λ_0), but whose second vector x_1 (the associated generalized eigenvector) is not. Adding x_1 to (X_0, Λ_0)

yields an enlarged invariant pair which is simple. In fact, the null space of the total derivative $\mathrm{D}\mathbf{F}$ at (X_0, Λ_0) provides all the necessary information to carry out this enlargement.

Corollary 4.1.8. *Let $(X_0, \Lambda_0) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{m\times m}$ be a minimal invariant pair of a real nonlinear eigenvalue problem and assume the conditions (J1) and (J2), and, equivalently, the conditions (TP1) and (TP2) to be satisfied. Then, the null space of $\mathrm{D}\mathbf{F}(X_0, \Lambda_0)$ is spanned by a pair $(\triangle X_0, \triangle\Lambda_0)$ having the form*

$$\triangle X_0 = y\psi^{\mathsf{T}}, \quad \triangle\Lambda_0 = z\psi^{\mathsf{T}}, \qquad y \in \mathbb{R}^n, \quad z, \psi \in \mathbb{R}^m, \tag{4.17}$$

where ψ is a left eigenvector belonging to the real eigenvalue μ of Λ_0 in (J2); i.e., $\psi^{\mathsf{T}}\Lambda_0 = \mu\psi^{\mathsf{T}}$. Furthermore, the extended pair

$$(\hat{X}_0,\ \hat{\Lambda}_0) = \left([X_0,\ y],\ \begin{bmatrix}\Lambda_0 & z\\ 0 & \mu\end{bmatrix}\right) \tag{4.18}$$

constitutes a simple invariant pair.

Proof. Theorem 4.1.3 implies that $\triangle X_0 = y\psi^{\mathsf{H}}$ and $\triangle\Lambda_0 = z\psi^{\mathsf{H}}$ with $y \in \mathbb{C}^{n\times r}$, $z, \psi \in \mathbb{C}^{m\times r}$, $\psi^{\mathsf{H}}\Lambda_0 = \mu\psi^{\mathsf{H}}$, $\mu \in \mathbb{C}^{r\times r}$ such that the extended pair $(\hat{X}_0, \hat{\Lambda}_0)$ in (4.18) is minimal invariant. The case $r = 0$, corresponding to $(\triangle X_0, \triangle\Lambda_0) = (0, 0)$, is excluded by (TP1), entailing $r \geq 1$. On the other hand, r cannot exceed 1 because the extended pair $(\hat{X}_0, \hat{\Lambda}_0)$ in (4.18) is minimal invariant and the multiplicity of (X_0, Λ_0) equals 2 by (J1). Hence, $r = 1$ and the extended pair $(\hat{X}_0, \hat{\Lambda}_0)$ in (4.18) is simple. Moreover, μ must be the real eigenvalue postulated by (J2), showing that the vectors y, z, ψ can be chosen real. □

4.2 A pseudo-arclength continuation algorithm

In the following, we consider a family of nonlinear eigenvalue problems (4.1), parameterized by a real, scalar parameter s. The dependence of the matrix-valued function T on s is assumed to be continuously differentiable.

Let (X_0, Λ_0) be a minimal invariant pair of the nonlinear eigenvalue problem (4.1) at a fixed parameter value $s = s_0$. The goal is now to continue this minimal invariant pair as the parameter s varies. By the considerations in Section 4.1, locally, the continuation of (X_0, Λ_0) as an invariant pair amounts to the continuation of (X_0, Λ_0) as a solution of the parameterized nonlinear equation

$$\mathbf{F}(X, \Lambda, s) = 0 \tag{4.19}$$

with $\mathbf{F}(X, \Lambda, s) = \left[\begin{smallmatrix}\mathbf{T}(X,\Lambda,s)\\ \mathbf{N}(X,\Lambda)\end{smallmatrix}\right]$, where $\mathbf{N}(\cdot,\cdot)$ is the normalization condition defined in (4.2) and $\mathbf{T}(\cdot,\cdot,s)$ is the block residual (3.4) associated with the matrix-valued function $T(\cdot, s)$.

4.2.1 Pseudo-arclength continuation. At first sight, it may appear natural to use the parameter s in the nonlinear eigenvalue problem (4.1) also for the continuation. This choice is reasonable if the invariant pair (X_0, Λ_0) is simple. In this case, we know from Theorem 4.1.2 that the Fréchet derivative $\mathrm{D}_{(X,\Lambda)}\mathbf{F}$ of $\mathbf{F}$ with respect to (X, Λ) is nonsingular at (X_0, Λ_0, s_0). Consequently, applying the Implicit Function Theorem to the nonlinear equation (4.19) yields a continuously differentiable dependence of (X, Λ) on s in the vicinity of s_0, and a simple path-following

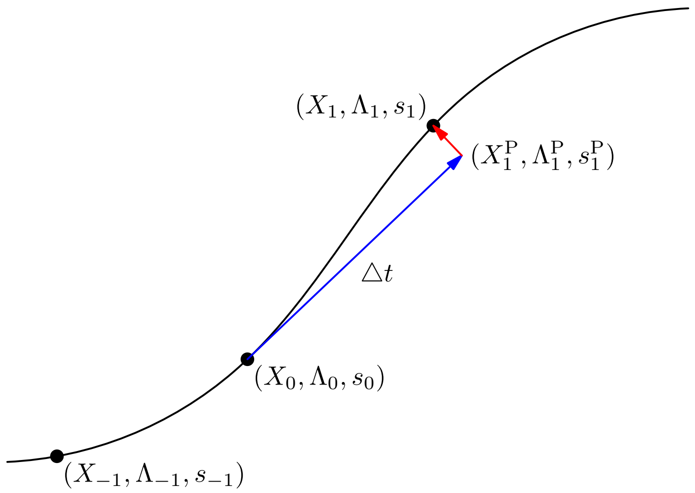
(X_1, Λ_1, s_1)
$(X_1^{\mathrm{P}}, \Lambda_1^{\mathrm{P}}, s_1^{\mathrm{P}})$
$\triangle t$
(X_0, Λ_0, s_0)
$(X_{-1}, \Lambda_{-1}, s_{-1})$

Figure 4.2: Illustration of a pseudo-arclength continuation algorithm.

algorithm is then sufficient to continue (X, Λ). However, if the aforementioned Fréchet derivative is singular, the parameterization of (X, Λ) by s might not be differentiable anymore or even break down.

Since non-simple invariant pairs caused by eigenvalue collisions (compare Section 4.1.2) can occur at any time during the continuation process, we need a more robust continuation algorithm which does not rely on the invertibilty of $\mathrm{D}_{(X,\Lambda)}\mathbf{F}$. In this work, we implement a standard pseudo-arclength continuation method; see, e.g., [50]. For this purpose, a reparameterization of the problem (4.19) is required: We now consider X, Λ, and s as being smoothly dependent on a new parameter t and look for a solution curve $\big(X(t), \Lambda(t), s(t)\big)$ such that

$$\mathbf{F}\big(X(t), \Lambda(t), s(t)\big) = 0.$$

Setting $\big(X(0), \Lambda(0), s(0)\big) = (X_0, \Lambda_0, s_0)$, the continuation of (X, Λ, s) with respect to t proceeds in two steps:

1. **Predictor.** The method computes the tangent to the solution curve at the current iterate and takes a step of length $\triangle t$ along the tangent.

2. **Corrector.** The point obtained in the first step is used as a starting value for Newton's method to find a nearby point on the solution curve, which then becomes the next iterate.

The procedure is visualized in Figure 4.2.

4.2.2 Predictor.

For ease of notation, we introduce the abbreviations

$$\mathrm{D}_X\mathbf{F}_0 = \mathbf{D}_X\mathbf{F}(X_0, \Lambda_0, s_0), \quad \mathrm{D}_\Lambda\mathbf{F}_0 = \mathrm{D}_\Lambda\mathbf{F}(X_0, \Lambda_0, s_0), \quad \mathrm{D}_s\mathbf{F}_0 = \mathrm{D}_s\mathbf{F}(X_0, \Lambda_0, s_0)$$

and denote derivatives with respect to t by dots. In order to determine the direction $(\dot{X}_0, \dot{\Lambda}_0, \dot{s}_0)$ of the tangent to the solution curve at the current iterate, we differentiate the nonlinear equation (4.19) with respect to t, resulting in the linear system

$$\mathrm{D}_X\mathbf{F}_0(\dot{X}_0) + \mathrm{D}_\Lambda\mathbf{F}_0(\dot{\Lambda}_0) + \mathrm{D}_s\mathbf{F}_0(\dot{s}_0) = 0.$$

Since the number of unknowns exceeds the number of equations by one, we have to impose an additional constraint. The normalization condition

$$\|\dot{X}_0\|^2 + \|\dot{\Lambda}_0\|^2 + \|\dot{s}_0\|^2 = 1$$

would be appropriate, but is nonlinear in the unknowns, and thus is replaced by

$$\langle \dot{X}_{-1}, \dot{X}_0 \rangle + \langle \dot{\Lambda}_{-1}, \dot{\Lambda}_0 \rangle + \langle \dot{s}_{-1}, \dot{s}_0 \rangle = 1.$$

where $(\dot{X}_{-1}, \dot{\Lambda}_{-1}, \dot{s}_{-1})$ is the tangential direction at the previous iterate. Whenever there is no previous iterate, we simply use $(\dot{X}_{-1}, \dot{\Lambda}_{-1}, \dot{s}_{-1}) = (0, 0, 1)$ to continue s in positive direction or $(\dot{X}_{-1}, \dot{\Lambda}_{-1}, \dot{s}_{-1}) = (0, 0, -1)$ to continue s in negative direction. The inner products are trace inner products, weighted by the number of entries, i. e.,

$$\langle \dot{X}_{-1}, \dot{X}_0 \rangle = \tfrac{1}{nm} \operatorname{tr} \dot{X}_{-1}^{\mathsf{T}} \dot{X}_0, \quad \langle \dot{\Lambda}_{-1}, \dot{\Lambda}_0 \rangle = \tfrac{1}{m^2} \operatorname{tr} \dot{\Lambda}_{-1}^{\mathsf{T}} \dot{\Lambda}_0, \quad \langle \dot{s}_{-1}, \dot{s}_0 \rangle = \dot{s}_{-1} \dot{s}_0. \tag{4.20}$$

In summary, we obtain a linear system for the computation of $(\dot{X}_0, \dot{\Lambda}_0, \dot{s}_0)$, which can be written in the block operator form

$$\begin{bmatrix} \mathrm{D}_X \mathbf{F}_0(\cdot) & \mathrm{D}_\Lambda \mathbf{F}_0(\cdot) & \mathrm{D}_s \mathbf{F}_0(\cdot) \\ \langle \dot{X}_{-1}, \cdot \rangle & \langle \dot{\Lambda}_{-1}, \cdot \rangle & \langle \dot{s}_{-1}, \cdot \rangle \end{bmatrix} \begin{bmatrix} \dot{X}_0 \\ \dot{\Lambda}_0 \\ \dot{s}_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{4.21}$$

This system can be shown to possess a unique solution even for non-simple pairs, where $\mathrm{D}_{(X,\Lambda)} \mathbf{F}_0 = \big[\mathrm{D}_X \mathbf{F}_0,\ \mathrm{D}_\Lambda \mathbf{F}_0\big]$ is singular, provided that the transversality condition

(TP3) $$\mathrm{D}_s \mathbf{F}_0 \notin \operatorname{im}\big[\mathrm{D}_X \mathbf{F}_0,\ \mathrm{D}_\Lambda \mathbf{F}_0\big],$$

the third condition characterizing a quadratic turning point after (TP1) and (TP2), is satisfied.

Once the tangential direction $(\dot{X}_0, \dot{\Lambda}_0, \dot{s}_0)$ of the solution curve has been computed from the linear system (4.21), a first-order prediction of the next iterate is given by

$$\big(X_1^{\mathrm{P}},\ \Lambda_1^{\mathrm{P}},\ s_1^{\mathrm{P}}\big) = \big(X_0,\ \Lambda_0,\ s_0\big) + \frac{\triangle t}{\eta} \big(\dot{X}_0,\ \dot{\Lambda}_0,\ \dot{s}_0\big), \tag{4.22}$$

where (X_0, Λ_0, s_0) is the current iterate and $\eta = \big[\langle \dot{X}_0, \dot{X}_0 \rangle + \langle \dot{\Lambda}_0, \dot{\Lambda}_0 \rangle + \langle \dot{s}_0, \dot{s}_0 \rangle\big]^{1/2}$.

4.2.3 Corrector. Based on the prediction (4.22) from the preceding subsection, the continued invariant pair $(X_1,\ \Lambda_1,\ s_1)$ is found by applying Newton's method to the nonlinear equation (4.19) with the predicted point $(X_1^{\mathrm{P}},\ \Lambda_1^{\mathrm{P}},\ s_1^{\mathrm{P}})$ as an initial guess. If the current iterate of the Newton procedure is denoted by $(X,\ \Lambda,\ s)$, the Newton correction $(\triangle X, \triangle \Lambda, \triangle s)$ has to satisfy

$$\mathrm{D}_X \mathbf{F}_0(\triangle X) + \mathrm{D}_\Lambda \mathbf{F}_0(\triangle \Lambda) + \mathrm{D}_s \mathbf{F}_0(\triangle s) = -\mathbf{F}(X, \Lambda, s).$$

Again, there is one more unknown than there are equations so that we have to impose an additional constraint to make the problem well posed. Therefore, we require that the correction be orthogonal to the tangential direction $(\dot{X}_0,\ \dot{\Lambda}_0,\ \dot{s}_0)$ computed in the prediction phase, i.e.,

$$\langle \dot{X}_0, \triangle X \rangle + \langle \dot{\Lambda}_0, \triangle \Lambda \rangle + \langle \dot{s}_0, \triangle s \rangle = 0,$$

where the inner products are again those defined in (4.20). Altogether, the Newton correction is defined as the solution to the linear system

$$\begin{bmatrix} \mathrm{D}_X\mathbf{F}_0(\cdot) & \mathrm{D}_\Lambda\mathbf{F}_0(\cdot) & \mathrm{D}_s\mathbf{F}_0(\cdot) \\ \langle \dot{X}_0, \cdot\rangle & \langle \dot{\Lambda}_0, \cdot\rangle & \langle \dot{s}_0, \cdot\rangle \end{bmatrix} \begin{bmatrix} \triangle X \\ \triangle\Lambda \\ \triangle s \end{bmatrix} = \begin{bmatrix} -\mathbf{F}(X,\Lambda,s) \\ 0 \end{bmatrix}.$$

4.2.4 Solving the linear systems. In both stages of the pseudo-arclength continuation algorithm, we are facing the need to solve linear systems involving derivatives of the nonlinear operator $\mathbf{F}$ together with some normalization or orthogonality condition. Recalling the definition of $\mathbf{F}$ at the beginning of Section 4.2, the systems in both cases turn out to be of the form

$$\begin{bmatrix} \mathrm{D}_X\mathbf{T}_0(\cdot) & \mathrm{D}_\Lambda\mathbf{T}_0(\cdot) & \mathrm{D}_s\mathbf{T}_0(\cdot) \\ \mathrm{D}_X\mathbf{N}_0(\cdot) & \mathrm{D}_\Lambda\mathbf{N}_0(\cdot) & \mathrm{D}_s\mathbf{N}_0(\cdot) \\ \langle \dot{X}, \cdot\rangle & \langle \dot{\Lambda}, \cdot\rangle & \langle \dot{s}, \cdot\rangle \end{bmatrix} \begin{bmatrix} \triangle X \\ \triangle\Lambda \\ \triangle s \end{bmatrix} = \begin{bmatrix} R \\ S \\ \triangle t \end{bmatrix}, \tag{4.23}$$

where 0-subscripts with derivatives again signify evaluation at $(X_0,\, \Lambda_0,\, s_0)$.

To solve the system (4.23), we first apply a unitary coordinate transformation which brings Λ_0 to (complex) Schur form. The structure of (4.23) is preserved under such a transformation, and we will assume, for simplicity, that (4.23) already refers to the transformed system.

It is easily seen from Lemma 3.3.8 that if Λ_0 is upper triangular, the j-th columns of $\mathrm{D}_X\mathbf{T}_0(\triangle X)$ and $\mathrm{D}_\Lambda\mathbf{T}_0(\triangle\Lambda)$ depend only on the first j columns of $\triangle X$ and $\triangle\Lambda$. The same is true for $\mathrm{D}_X\mathbf{N}_0(\triangle X)$ and $\mathrm{D}_\Lambda\mathbf{N}_0(\triangle\Lambda)$ by an analogous consideration. In other words, for suitable linear operators $\left[\mathrm{D}_X\mathbf{T}_0\right]_{ij}$, $\left[\mathrm{D}_\Lambda\mathbf{T}_0\right]_{ij}$, $\left[\mathrm{D}_X\mathbf{N}_0\right]_{ij}$, and $\left[\mathrm{D}_\Lambda\mathbf{N}_0\right]_{ij}$,

$$\begin{aligned} \left[\mathrm{D}_X\mathbf{T}_0(\triangle X)\right]_j &= \sum_{i=1}^{j}\left[\mathrm{D}_X\mathbf{T}_0\right]_{ij}\triangle X_i, & \left[\mathrm{D}_\Lambda\mathbf{T}_0(\triangle\Lambda)\right]_j &= \sum_{i=1}^{j}\left[\mathrm{D}_\Lambda\mathbf{T}_0\right]_{ij}\triangle\Lambda_i, \\ \left[\mathrm{D}_X\mathbf{N}_0(\triangle X)\right]_j &= \sum_{i=1}^{j}\left[\mathrm{D}_X\mathbf{N}_0\right]_{ij}\triangle X_i, & \left[\mathrm{D}_\Lambda\mathbf{N}_0(\triangle\Lambda)\right]_j &= \sum_{i=1}^{j}\left[\mathrm{D}_\Lambda\mathbf{N}_0\right]_{ij}\triangle\Lambda_i. \end{aligned}$$

This fact suggests a columnwise forward substitution scheme to solve (4.23).

More specifically, we will adapt the bordered Bartels-Stewart algorithm from [23, 83] to the general nonlinear case. Using the notation introduced above, we define the matrices

$$L_j = \begin{bmatrix} \left[\mathrm{D}_X\mathbf{T}_0\right]_{jj} & \left[\mathrm{D}_\Lambda\mathbf{T}_0\right]_{jj} & \mathrm{D}_s\mathbf{T}_0 \\ \left[\mathrm{D}_X\mathbf{N}_0\right]_{jj} & \left[\mathrm{D}_\Lambda\mathbf{N}_0\right]_{jj} & 0 \\ \frac{1}{nm}\dot{X}_j & \frac{1}{m^2}\dot{\Lambda}_j & \frac{1}{m}\dot{s}_0 \end{bmatrix}$$

and solve the linear systems

$$L_j \begin{bmatrix} \triangle X_j^0 \\ \triangle \Lambda_j^0 \\ \triangle s_j^0 \end{bmatrix} = \begin{bmatrix} R_j - \sum\limits_{i=1}^{j-1} \Big(\left[\mathrm{D}_X \mathbf{T}_0 \right]_{ij} \triangle X_i^0 + \left[\mathrm{D}_\Lambda \mathbf{T}_0 \right]_{ij} \triangle \Lambda_i^0 \Big) \\ S_j - \sum\limits_{i=1}^{j-1} \Big(\left[\mathrm{D}_X \mathbf{N}_0 \right]_{ij} \triangle X_i^0 + \left[\mathrm{D}_\Lambda \mathbf{N}_0 \right]_{ij} \triangle \Lambda_i^0 \Big) \\ 0 \end{bmatrix},$$

$$L_j \begin{bmatrix} \triangle X_j^j \\ \triangle \Lambda_j^j \\ \triangle s_j^j \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

as well as

$$L_j \begin{bmatrix} \triangle X_j^k \\ \triangle \Lambda_j^k \\ \triangle s_j^k \end{bmatrix} = \begin{bmatrix} -\sum\limits_{i=k}^{j-1} \Big(\left[\mathrm{D}_X \mathbf{T}_0 \right]_{ij} \triangle X_i^k + \left[\mathrm{D}_\Lambda \mathbf{T}_0 \right]_{ij} \triangle \Lambda_i^k \Big) \\ -\sum\limits_{i=k}^{j-1} \Big(\left[\mathrm{D}_\Lambda \mathbf{N}_0 \right]_{ij} \triangle X_i^k + \left[\mathrm{D}_\Lambda \mathbf{N}_0 \right]_{ij} \triangle \Lambda_i^k \Big) \\ 0 \end{bmatrix}, \qquad k = 1, \dots, j-1$$

for $j = 1, \dots, m$. The total number of systems to be solved is $\frac{1}{2}m(m+3)$. The j-th column of the solution to the linear system (4.23) is then given by the linear combination

$$\triangle X_j = \triangle X_j^0 + \sum_{k=1}^{j} \alpha_k \triangle X_j^k, \quad \triangle \Lambda_j = \triangle \Lambda_j^0 + \sum_{k=1}^{j} \alpha_k \triangle \Lambda_j^k, \quad \triangle s = \triangle s_1^0 + \alpha_1 \triangle s_1^1,$$

where the coefficients $\alpha_1, \dots, \alpha_m$ satisfy

$$\begin{bmatrix} \triangle s_2^1 - \triangle s_1^1 & \triangle s_2^2 & & \\ \vdots & \vdots & \ddots & \\ \triangle s_m^1 - \triangle s_1^1 & \triangle s_m^2 & \cdots & \triangle s_m^m \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} = \begin{bmatrix} \triangle s_1^0 - \triangle s_2^0 \\ \vdots \\ \triangle s_1^0 - \triangle s_m^0 \\ \triangle t \end{bmatrix}.$$

4.2.5 Step size control. It may happen that the Newton corrector outlined in Section 4.2.3 fails to converge or convergence is very slow. In such situations, it is natural to terminate the Newton process after a prescribed number of steps and to perform another prediction with a reduced step size. The hope is that the smaller step size will lead to a more accurate prediction from the first-order model and, through this, cause the convergence behavior to improve. In our implementation, we cut the step size in half if the Newton procedure does not converge within 5 iterations.

On the other hand, it may also be beneficial to increase the step length if convergence occurs rapidly. Of course, this is likely to decelerate the convergence of the corrector steps a little. But at the same time it will diminish the total number of predictor-corrector cycles needed to traverse a given range of parameters, so we may still save some work. We enlarge the step by 10% in our implementation if 3 or less Newton iterations are necessary until convergence. For more than 3 but not more than 5 iterations, the step length is left unchanged. A similar strategy for step size control has been pursued in [23] for quadratic eigenvalue problems.

4.2.6 Turning points. As already discussed in Section 4.2.1, the pseudo-arclength continuation approach is robust and reliable even in the presence of quadratic turning points characterized by the conditions (TP1), (TP2), and (TP3). However, it is well known [50] that at a quadratic turning point, $\dot{s}$ switches signs, which corresponds to a reversal of the direction of s at the turning point, hence the name. Since we want the parameter s to run through an interval, this behavior is undesirable, and we would rather like the direction of s to remain the same.

4.2.7 Augmenting a non-simple invariant pair. Recall that according to Theorem 4.1.6, a quadratic turning point occurs if and only if a real eigenvalue inside the invariant pair being continued collides with another real eigenvalue from the outside, forming a complex conjugate pair. Obviously, in real arithmetic, it is not possible to follow only one of the ensuing complex eigenvalues without taking into account the other one as well. Consequently, we have to include the missing eigenvalue in the current invariant pair to be able to carry on the continuation. Corollary 4.1.8 provides the foundations for doing so.

Suppose the continuation procedure outlined in Section 4.2.1 has reached a quadratic turning point (X_*, Λ_*, s_*). Details on how turning points are detected and computed will be given in the next subsection. Then, the $\dot{s}$-component of the corresponding tangential direction $(\dot{X}_*, \dot{\Lambda}_*, \dot{s}_*)$ as determined from the linear system (4.21) will be zero. Consequently, $(\dot{X}_*, \dot{\Lambda}_*)$ spans the null space of $\mathrm{D}_{(X,\Lambda)}\mathbf{F}$ at (X_*, Λ_*, s_*). Thus, $\dot{\Lambda}_*$ is a rank-one matrix of the form (4.17) by Corollary 4.1.8, and we may obtain an extended invariant pair, which is simple, via the update

$$(\hat{X}_*, \hat{\Lambda}_*) = \left(\begin{bmatrix} X_*, & \dot{X}_* v_1 \end{bmatrix}, \begin{bmatrix} \Lambda_* & u_1 \sigma_1 \\ 0 & v_1^T \Lambda_* v_1 \end{bmatrix} \right),$$

where σ_1 is the largest singular value of $\dot{\Lambda}_*$ and u_1, v_1 are corresponding left and right singular vectors. After the update, a few steps of Newton's iteration should be executed starting from $(\hat{X}_*, \hat{\Lambda}_*, s_*)$ to make sure that the new pair is truly invariant also in the presence of round-off error.

4.2.8 Detecting and computing quadratic turning points. In order to identify the occurence of quadratic turning points during the continuation process, we monitor the derivatives $\dot{s}$ computed in the prediction stage of every iteration for sign changes. As long as the sign of $\dot{s}$ remains unchanged, so does the step direction of s, and we can keep on iterating in the usual manner. If, on the other hand, the derivatives $\dot{s}(t_0)$ and $\dot{s}(t_1)$ in two consecutive iterations with t-values t_0 and t_1, respectively, are found to have opposite signs, the Intermediate Value Theorem ensures the existence of a $t_* \in (t_0, t_1)$ such that $\dot{s}(t_*) = 0$. That is, a potential turning point $(X_*,\ \Lambda_*,\ s_*) = \big(X(t_*),\ \Lambda(t_*),\ s(t_*)\big)$ has been detected.

To compute t_*, we proceed as described in [35, pp. 259–261]. First, a cubic Hermite interpolating polynomial is constructed which matches both the values and the first derivatives of the curve $\big(X(t),\ \Lambda(t),\ s(t)\big)$ at $t = t_0$ as well as $t = t_1$. This yields a first estimate of the turning point by choosing the value $\big(X_*^{\mathrm{P}},\ \Lambda_*^{\mathrm{P}},\ s_*^{\mathrm{P}}\big)$ of the polynomial at the point $t_*^{\mathrm{P}} \in (t_0, t_1)$ for which s_*^{P} attains an extremum; see Figure 4.3 for an illustration. In most cases, this estimate or, more precisely, the derivative $\big(\dot{X}_*^{\mathrm{P}},\ \dot{\Lambda}_*^{\mathrm{P}},\ \dot{s}_*^{\mathrm{P}}\big)$ at t_*^{P} predicted by the polynomial will be enough to carry out the enlargement of the invariant pair outlined in Section 4.2.7. In fact, this has been the case in all of our experiments. In the rare event that this estimate is not sufficiently accurate, it can be further refined by a bisection approach with

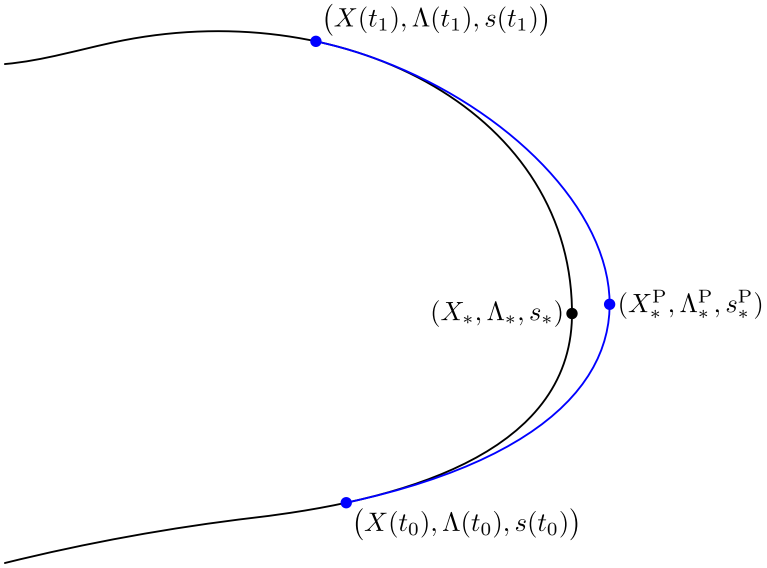
$(X(t_1), \Lambda(t_1), s(t_1))$
(X_*, Λ_*, s_*)
$(X_*^{\mathrm{P}}, \Lambda_*^{\mathrm{P}}, s_*^{\mathrm{P}})$
$(X(t_0), \Lambda(t_0), s(t_0))$

Figure 4.3: An estimate $(X_*^{\mathrm{P}}, \Lambda_*^{\mathrm{P}}, s_*^{\mathrm{P}})$ of the turning point is obtained from the extremal point of a cubic interpolating polynomial (blue).

respect to t. To this end, the estimate is corrected back to the solution curve by the Newton corrector in Section 4.2.3, and the procedure is repeated with either t_0 or t_1 replaced by t_*^{P}.

4.2.9 Overall algorithm. If we put all the individual parts from the previous subsections together, we obtain Algorithm 4.1 for continuing a minimal invariant pair for the parameter s running from s_0 into positive direction. To have s running in negative direction, substitute $\dot{s}_0 := -1$ for $\dot{s}_0 := 1$ in the initialization stage of the algorithm.

4.3 Numerical experiments

To verify our implementation of the numerical continuation method detailed in Section 4.2, we have applied it to two academic test problems.

4.3.1 A synthetic test case. As our first example, we consider the nonlinear eigenvalue problem

$$\big(\lambda I - A_0(s) - \mathrm{e}^{-\lambda} A_1(s)\big)x = 0, \quad x \neq 0, \tag{4.24}$$

where the entries of the coefficient matrices A_0 and A_1 depend smoothly on the real parameter s. Nonlinear eigenvalue problems of this type arise, for example, in the stability analysis of delay differential equations with a single constant delay; see Section 2.1.

For the sake of simplicity, we choose the coefficient matrices A_0 and A_1 to be diagonal. More specifically, we set

$$A_0(s) = \operatorname{diag}\{2, 3, \dots, 7\} - sI, \qquad A_1(s) = -I.$$

Algorithm 4.1: Continuation of minimal invariant pairs for nonlinear eigenvalue problems

```
Input: block residual $(X, \Lambda, s) \mapsto \mathbf{T}(X, \Lambda, s)$, initial parameter $s_0$,
       (approximate) initial minimal invariant pair $(X_0^{\mathrm{P}}, \Lambda_0^{\mathrm{P}})$ at $s_0^{\mathrm{P}} = s_0$,
       initial step size $\triangle t$, maximal expected minimality index $\ell$
Output: continued minimal invariant pairs $(X_i, \Lambda_i)$ for $i = 0, 1, 2, \dots$ at
        parameter values $s_0 < s_1 < s_2 < \cdots$
% Initialization
$\dot{X}_{-1} := 0, \qquad \dot{\Lambda}_{-1} := 0, \qquad \dot{s}_{-1} := 1$
$W := \mathbf{V}_\ell(X_0^{\mathrm{P}}, \Lambda_0^{\mathrm{P}})$
% Continuation
for $i = 0, 1, \dots$ do
    % Corrector
    Apply Newton method from Section 4.2.3 to obtain minimal invariant
    pair $(X_i, \Lambda_i)$ at parameter value $s_i$ from estimate $(X_i^{\mathrm{P}}, \Lambda_i^{\mathrm{P}}, s_i^{\mathrm{P}})$.
    if Newton process does not converge then
        Reduce step size $\triangle t$. Return to predictor if sensible and terminate
        otherwise.
    end
    % Predictor
    Update $W := \mathbf{V}_\ell(X_i, \Lambda_i)$.
    Compute tangential direction $(\dot{X}_i, \dot{\Lambda}_i, \dot{s}_i)$ at $(X_i, \Lambda_i, s_i)$ from (4.21).
    Handling of turning points
    if $\dot{s}_i \cdot \dot{s}_{i-1} < 0$ then
        Compute turning point as described in Section 4.2.8.
        Augment invariant pair according to Section 4.2.7 and store result in
        $(X_{i+1}^{\mathrm{P}}, \Lambda_{i+1}^{\mathrm{P}}, s_{i+1}^{\mathrm{P}})$.
    else
        Determine $(X_{i+1}^{\mathrm{P}}, \Lambda_{i+1}^{\mathrm{P}}, s_{i+1}^{\mathrm{P}})$ by taking a step of length $\triangle t$ along the
        computed tangent.
    end
end
```

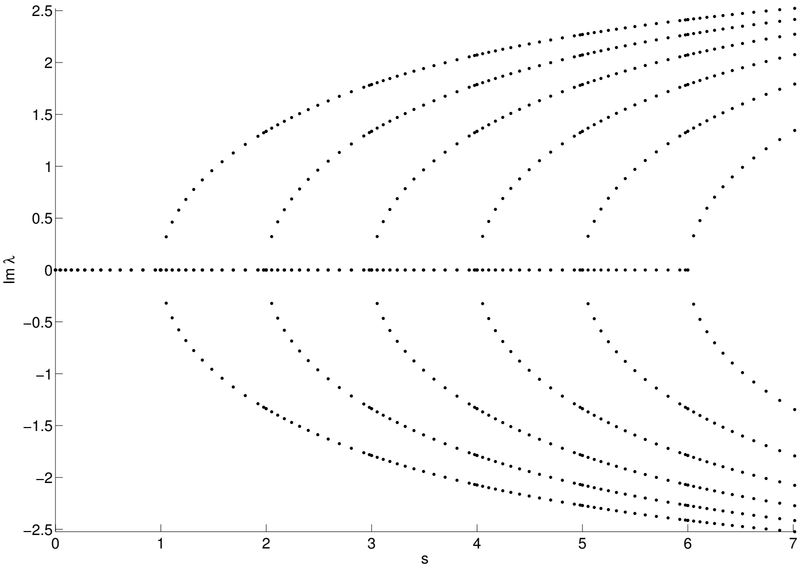

Figure 4.4: Imaginary parts of the eigenvalues inside the minimal invariant pair being continued as the parameter s varies from 0 to 7.

With this choice, the characteristic equation of the nonlinear eigenvalue problem (4.24) becomes

$$(\lambda + \mathrm{e}^{-\lambda} + s - 2) \cdots (\lambda + \mathrm{e}^{-\lambda} + s - 7) = 0. \tag{4.25}$$

From the monotonicity behavior of the mapping $\lambda \mapsto \lambda + \mathrm{e}^{-\lambda}$, it is clear that the i-th factor on the left-hand side of the characteristic equation (4.25) has

- two real roots for $s < i$,
- one real root $\lambda = 0$ of multiplicity 2 for $s = i$,
- no real root for $s > i$.

Consequently, for the parameter value $s = 0$, the nonlinear eigenvalue problem (4.24) has 12 real eigenvalues. If the parameter s is now raised to 1, two of these real eigenvalues collide and form a complex conjugate pair. At $s = 2$, the same happens for the next pair of real eigenvalues and so on until, finally, the last pair of real eigenvalues coalesces at $s = 6$. For even higher values of the parameter s, all eigenvalues of (4.24) are complex.

By solving the linear eigenvalue problem which remains after neglecting the exponential term in (4.24), we find the approximate minimal invariant pair

$$\big(X_0,\ \Lambda_0\big) = \big(I,\ \mathrm{diag}\{2, \ldots, 7\}\big)$$

for the parameter value $s = 0$. We pass this minimal invariant pair to our code and ask for continuation up to $s = 7$. As it turns out, this (approximate) initial minimal invariant pair contains precisely one eigenvalue from each of the colliding eigenvalue pairs mentioned previously and one can clearly see from Figure 4.4 that one pair of complex conjugate eigenvalues emerges for every $s = 1, \dots, 6$.

Another point worth noting about this problem is that the pair of eigenvalues originating from the i-th factor on the left-hand side of the characteristic equation (4.25) shares the same eigenvector, namely the i-th unit vector. Therefore, the minimality index of the resulting minimal invariant pair at $s = 7$ is two. Hence, we have to set $\ell = 2$ in the normalization condition (4.2).

4.3.2 Parabolic PDE with delay. We consider a parabolic partial differential equation with a time delay τ:

$$\frac{\partial u}{\partial t}(x,t) = \frac{\partial^2 u}{\partial x^2}(x,t) + a_0 \cdot u(x,t) + a_1(x) \cdot u(x, t-\tau)$$

$$u(0,t) = u(\pi,t) = 0$$

with $a_0 = 20$, $a_1(x) = -4.1 + x(1 - \mathrm{e}^{x-\pi})$. This example is taken from [65, Sec. 2.4.1], which, in turn, is a modification of [142, Chapter 3, Example 1.12] in that it allows the coefficient a_1 to be dependent on the spatial variable x. A spatial discretization by finite differences with the uniform grid size $h = \frac{\pi}{n+1}$ yields the delay differential equation

$$\dot{v}(t) = A_0 v(t) + A_1 v(t-\tau) \tag{4.26}$$

of dimension n, where $v(t) = \big[u(x_1,t), \dots, u(x_n,t)\big]^{\mathsf{T}}$ with $x_i = \frac{i}{n+1}\pi$, $i = 1, \dots, n$, and the coefficient matrices $A_0, A_1 \in \mathbb{R}^{n \times n}$ are given by

$$A_0 = \Big(\frac{n+1}{\pi}\Big)^2 \begin{bmatrix} -2 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & -2 \end{bmatrix} + a_0 I, \quad A_1 = \begin{bmatrix} a_1(x_1) & & \\ & \ddots & \\ & & a_1(x_n) \end{bmatrix}.$$

For the stability analysis of the delay differential equation (4.26), one is interested in a few eigenvalues with largest real part of the nonlinear eigenvalue problem

$$\big(-\lambda I + A_0 + \mathrm{e}^{-\tau\lambda} A_1\big) v = 0, \tag{4.27}$$

which depends on the delay τ as a parameter. In the special case $\tau = 0$, i.e., when there is no delay, the eigenvalue problem (4.27) is, in fact, linear and symmetric. Therefore, its eigenvalues can be easily computed by standard methods and turn out to be all real. When increasing the delay, several eigenvalues remain real while others collide and form complex conjugate pairs. We apply our continuation algorithm for $n = 100$ to the six eigenvalues with largest real part at $\tau = 0$ and continue them until $\tau = 0.4$. On two occasions eigenvalues collide. The first collision takes place at $\tau \approx 0.051$ and the other one at $\tau \approx 0.078$. In both cases, the step size is decreased and the invariant pair is enlarged. Figure 4.5 illustrates the obtained results.

Contributions within this chapter

In this chapter, we have developed a pseudo-arclength continuation algorithm for minimal invariant pairs of a real nonlinear eigenvalue problem depending on a

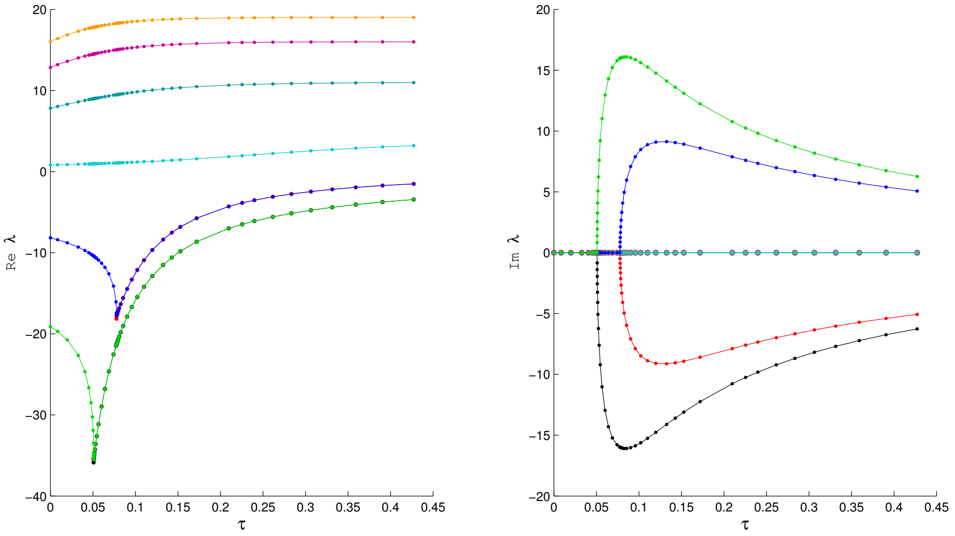

Figure 4.5: Continued eigenvalues vs. delay τ for the delay eigenvalue problem (4.27). Left: real part. Right: imaginary part.

single real design parameter. Special attention has been paid to the treatment of generic bifurcations that might be encountered during the continuation process. The resulting algorithm has been tested for two examples related to delay eigenvalue problems.

In Section 4.1, we establish the foundations for the continuation algorithm by characterizing minimal invariant pairs as zeros of a certain nonlinear function. Lemma 4.1.1 shows that this representation is invariant with respect to similarity transformations of the minimal invariant pair. In Theorem 4.1.2, we give a novel proof of the fact that the Fréchet derivative of the nonlinear function is invertible at the minimal invariant pair if and only if the pair is simple, which is shorter and more direct than the original proof in [83, Theorem 10]. Furthermore, we characterize the null space of the derivative for non-simple invariant pairs: A new result in Theorem 4.1.3 demonstrates a correspondence between the elements of the null space and possible augmentations of the non-simple invariant pair. This link fosters a direct proof for the central result of this section in Theorem 4.1.6, stating that, for real nonlinear eigenvalue problems, bifurcations generically occur only upon collision of two eigenvalues on the real axis. The original proof for this theorem by Beyn, Kressner and the author in [21] proceeded in several steps, showing the statement first for generalized linear eigenvalue problem and then extending it via polynomial to general holomorphic eigenvalue problems. The proof presented here is unpublished work by the author and leads straight to the general statement. In Corollary 4.1.8, it is shown how bifurcations can be cured by adding the incoming eigenvalue to the minimal invariant pair being continued. Also for this statement, we give a new proof, which is considerably shorter and simpler than the original proof due to Beyn, Kressner, and the author in [21].

Section 4.2 describes the various algorithmic details of continuation algorithm.

Most notably, we establish a bordered Bartels-Stewart algorithm analogous to the ones for linear and quadratic eigenvalue problems in [22] and [23], respectively, for the general holomorphic setting in Section 4.2.4. Moreover, in Section 4.2.7, we present an augmentation strategy for the minimal invariant pair, which is slightly modified in comparison to the ones in [22] and [23] and tends to achieve a better numerical stability in certain situations.

Chapter 5

Interpolation-based solution of nonlinear eigenvalue problems

This chapter focuses on nonlinear eigenvalue problems (1.1) for which evaluating the matrix-valued function T at some $\lambda \in \mathbb{C}$ is very expensive. A prime example for this situation are nonlinear eigenvalue problems stemming from boundary-element discretizations of operator eigenvalue problems as described in Section 2.3. For these problems, every entry of the matrix $T(\lambda)$ is defined in terms of a double boundary integral with a nonlocal and singular kernel function, leading to a far higher cost of assembly than for matrices originating from typical finite-element discretizations.

Most existing solution techniques for nonlinear eigenvalue problems prove inefficient for this type of problems since they are not designed to keep the number of function evaluations low. For instance, many methods rely on frequent formations of the residual $T(\tilde{\lambda})\tilde{x}$ for approximate eigenpairs $(\tilde{x}, \tilde{\lambda})$. A possible exception are methods based on contour integrals; see Section 1.2.2.

The approach considered in this chapter performs all necessary evaluations of the matrix-valued function T in a preprocessing step to construct a matrix-valued polynomial P approximating T in the region of interesting eigenvalues. Any further calculations are then carried out using P instead of T. This has two advantages: (i) evaluating P is considerably cheaper than evaluating T; (ii) the eigenvalue problem for P can be solved by a standard linear eigensolver after an appropriate linearization [48, 93]. Similar approaches have been proposed in [69, 130, 25, 72].

To assess the loss of accuracy incurred by the polynomial approximation, a first-order perturbation analysis for nonlinear eigenvalue problems is performed in Section 5.2. Combined with an approximation result for Chebyshev interpolation, this shows exponential convergence of the obtained eigenvalue approximations with respect to the degree of the approximating polynomial P.

5.1 Derivation of the method

5.1.1 Polynomial approximation. In the following, we constrain ourselves to the situation that the eigenvalue region of interest is the interval $[-1, 1]$. This covers general finite intervals or even prescribed smooth curves in the complex plane through an appropriate analytic reparameterization.

The main idea of our approach for solving the nonlinear eigenvalue problem (1.1) is to replace T by a polynomial approximant P. More specifically, for fixed interpolation nodes $\lambda_0, \lambda_1, \ldots, \lambda_d \in [-1,1]$, we replace T by the unique matrix-valued polynomial P of degree at most d satisfying the interpolation conditions

$$P(\lambda_j) = T(\lambda_j), \qquad j = 0, \ldots, d. \tag{5.1}$$

This leads to the polynomial eigenvalue problem

$$P(\lambda)x = 0. \tag{5.2}$$

We expect that a small interpolation error will lead to a small error in the eigenpairs. This expectation is confirmed by an error analysis in Section 5.2. Standard choices of interpolation nodes include Chebyshev nodes of the first kind,

$$\lambda_j = \cos\left(\frac{j+\frac{1}{2}}{d+1}\pi\right), \qquad j = 0, \ldots, d, \tag{5.3}$$

and of the second kind,

$$\lambda_j = \cos\left(\frac{j}{d}\pi\right), \qquad j = 0, \ldots, d. \tag{5.4}$$

As is well known and shown for our particular situation in Proposition 5.2.6 below, the interpolation error of such a Chebyshev interpolant decays exponentially with d. Hence, we expect that a moderate polynomial degree will be sufficient to ensure good accuracy.

5.1.2 Linearization of the polynomial eigenproblem. Having substituted the interpolating polynomial P for the nonlinear function T, we are facing the need to solve the resulting polynomial eigenvalue problem (5.2). A popular way of solving polynomial eigenvalue problems is to transform them into an equivalent (generalized) linear eigenvalue problem and then apply standard techniques. This transformation is not at all unique [93]. A common choice are companion linearizations based on an expansion of the polynomial P in the monomial basis. However, there is a number of inconveniences associated with the use of the monomial basis. First of all, the coefficient matrices of P with respect to the monomial basis are not readily available from the construction in Section 5.1.1. Moreover, especially for higher degrees of P, this transformation may cause numerical difficulties. Therefore, we employ a different linearization scheme described in [3], which is based on an expansion of P in the polynomial basis formed by the first $d+1$ Chebyshev polynomials (of the first kind),

$$P(\lambda) = P_0\tau_0(\lambda) + \cdots + P_d\tau_d(\lambda). \tag{5.5}$$

Combining the expansion (5.5) with the interpolation conditions (5.1), through which P is defined, leads to

$$T(\lambda_j) = \sum_{i=0}^{d} P_i \cos\frac{i(j+\frac{1}{2})\pi}{d+1}, \qquad j = 0, \ldots, d$$

if the Chebyshev nodes of the first kind in (5.3) are used as interpolation nodes, and to

$$T(\lambda_j) = \sum_{i=0}^{d} P_i \cos \frac{ij\pi}{d}, \qquad j = 0, \dots, d$$

for the Chebyshev nodes of the second kind in (5.4). In both cases, the coefficient matrices $P_0, \dots, P_d$ can be efficiently computed by a sequence of inverse discrete cosine transforms of type III or type I, respectively. For details, the reader is referred to, e.g., [12].

For the sake of completeness, let us recall the linearization technique from [3] for the polynomial eigenvalue problem

$$\big(P_0\tau_0(\lambda) + \cdots + P_d\tau_d(\lambda)\big)x = 0 \tag{5.6}$$

expressed in the Chebyshev basis. Introducing the vectors $x_k := \tau_k(\lambda)x$ for $k = 0, \dots, d$, the polynomial eigenvalue problem (5.6) can be rewritten as

$$P_0x_0 + \cdots + P_dx_d = 0. \tag{5.7}$$

Furthermore, the three-term recurrence for the Chebyshev polynomials τ_k yields

$$x_1 = \lambda x_0 \qquad \text{and} \qquad x_k = 2\lambda x_{k-1} - x_{k-2}, \quad k = 2, \dots, d.$$

By means of the preceding identities, we can eliminate x_d from the polynomial eigenvalue problem (5.7). The remaining equation,

$$P_0x_0 + \cdots + P_{d-3}x_{d-3} + (P_{d-2} - P_d)x_{d-2} + P_{d-1}x_{d-1} + 2\lambda P_dx_{d-1} = 0,$$

can be reformulated as the equivalent (generalized) linear eigenvalue problem

$$\mathcal{L}_0y = \lambda\mathcal{L}_1y \tag{5.8}$$

with $y = [x_0^\mathsf{T}, \dots, x_{d-1}^\mathsf{T}]^\mathsf{T}$ and

$$\mathcal{L}_0 = \begin{bmatrix} 0 & I & & & \\ I & 0 & I & & \\ & \ddots & \ddots & \ddots & \\ & & I & 0 & I \\ -P_0 & \cdots & -P_{d-3} & P_d - P_{d-2} & -P_{d-1} \end{bmatrix}, \quad \mathcal{L}_1 = \begin{bmatrix} I & & & & \\ & 2I & & & \\ & & \ddots & & \\ & & & 2I & \\ & & & & 2P_d \end{bmatrix}. \tag{5.9}$$

It has been shown in [3] that (5.8)–(5.9) is a strong linearization of the polynomial eigenvalue problem (5.6).

5.1.3 Solution of the linearized eigenproblem.

The resulting linearizations in (5.8)–(5.9) are typically large. Their size is equal to the size of the original nonlinear eigenvalue problem times the degree of the interpolating polynomial P. The eigenvalues of interest are those lying in or close to the real interval $[-1, 1]$. As these are likely to be interior eigenvalues of the problem, we pursue a shift-and-invert strategy for their computation. A natural choice for the shift is the center of the interval, i.e., zero. This choice leads us to the computation of a few eigenvalues of largest magnitude for the matrix $\Phi = \mathcal{L}_0^{-1}\mathcal{L}_1$, which can be easily accomplished using Krylov subspace methods, such as the implicitly restarted Arnoldi algorithm [87].

Krylov subspace methods crucially depend on repeated matrix-vector multiplication with the matrix Φ, which, in our case, can be broken up into successive multiplications by $\mathcal{L}_1$ and $\mathcal{L}_0^{-1}$. Whereas the multiplication by the block diagonal matrix $\mathcal{L}_1$ can be performed efficiently in a straightforward manner, the question of how to invert $\mathcal{L}_0$ is more subtle and will be treated subsequently.

The linear system $\mathcal{L}_0 u = v$ has the block structure

$$\begin{bmatrix} 0 & I & & & \\ I & 0 & I & & \\ & \ddots & \ddots & \ddots & \\ & & I & 0 & I \\ -P_0 & \cdots & -P_{d-3} & P_d - P_{d-2} & -P_{d-1} \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{d-2} \\ u_{d-1} \end{bmatrix} = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{d-2} \\ v_{d-1} \end{bmatrix}, \tag{5.10}$$

where we have partitioned the vectors u and v in accordance with $\mathcal{L}_0$. The odd-numbered block rows of (5.10) amount to the recursion

$$u_1 = v_0, \qquad u_{2j+1} = v_{2j} - u_{2j-1}, \quad j = 1, 2, 3, \dots,$$

which permits us to compute the entries u_1, u_3, u_5, ... of the solution. In a similar fashion, the even-numbered block rows give

$$u_{2j} = \hat{v}_{2j-1} + (-1)^j u_0, \quad j = 1, 2, 3, \dots, \tag{5.11}$$

where the vectors $\hat{v}_{2j-1}$ are determined by the recurrence

$$\hat{v}_1 = v_1, \qquad \hat{v}_{2j+1} = v_{2j+1} - \hat{v}_{2j-1}, \quad j = 1, 2, 3, \dots .$$

Inserting identity (5.11) into the last block row of the linear system (5.10), we arrive at the linear equation

$$\begin{aligned} (-P_0 + P_2 - P_4 + P_6 - \cdots + \cdots) u_0 = v_{d-1} &- P_d v_{d-1} \\ &+ (P_1 u_1 + P_3 u_3 + P_5 u_5 + \cdots) \\ &+ (P_2 \hat{v}_1 + P_4 \hat{v}_3 + P_6 \hat{v}_5 + \cdots), \end{aligned}$$

which needs to be solved for u_0. The system matrix

$$(-P_0 + P_2 - P_4 + P_6 - \cdots + \cdots)$$

should be LU factorized once in a preprocessing step before the actual Krylov subspace method is invoked. In this way, each application of $\mathcal{L}_0^{-1}$ requires only one pair of forward and backward solves. After u_0 has been computed, the remaining components u_2, u_4, u_6, ... are determined via (5.11).

5.1.4 Extraction of minimal invariant pairs. An invariant pair (X, Λ) of the matrix polynomial (5.5) satisfies

$$P_0 X \tau_0(\Lambda) + \cdots + P_d X \tau_d(\Lambda) = 0. \tag{5.12}$$

In the following, we will show how such an invariant pair can be obtained from the corresponding linearization $\mathcal{L}_0 - \lambda\mathcal{L}_1$ defined in (5.8)–(5.9). A linear eigensolver, such as the Arnoldi method discussed in Section 5.1.3, applied to $\mathcal{L}_0 - \lambda\mathcal{L}_1$ yields an invariant pair $(Y, \Lambda) \in \mathbb{C}^{dn \times m} \times \mathbb{C}^{m \times m}$ of the linearized problem, characterized by the relation

$$\mathcal{L}_0 Y = \mathcal{L}_1 Y \Lambda. \tag{5.13}$$

Note that $\operatorname{span}(Y)$ is usually called an invariant subspace of the matrix pencil $\mathcal{L}_0 - \lambda\mathcal{L}_1$.

Partitioning $Y = [X_0^\mathsf{T}, \dots, X_{d-1}^\mathsf{T}]^\mathsf{T}$ with $X_j \in \mathbb{C}^{n\times m}$ and exploiting the block structure of $\mathcal{L}_0, \mathcal{L}_1$, the first $d-1$ block rows of (5.13) amount to

$$X_1 = X_0\Lambda, \qquad X_{k-2} + X_k = 2X_{k-1}\Lambda, \quad k = 2, \dots, d-1. \tag{5.14}$$

A simple induction employing the three-term recurrence for the Chebyshev polynomials τ_k shows that (5.14) implies $X_k = X_0\tau_k(\Lambda)$ for $k = 0, \dots, d-1$. Inserting these relations into the last block row of (5.13), we find

$$-\sum_{k=0}^{d-1} P_k X_0 \tau_k(\Lambda) + P_d X_0 \tau_{d-2}(\Lambda) = 2P_d X_0 \tau_{d-1}(\Lambda)\Lambda,$$

and rearranging terms yields

$$-\sum_{k=0}^{d-1} P_k X_0 \tau_k(\Lambda) = P_d X_0 \big(2\tau_{d-1}(\Lambda)\Lambda - \tau_{d-2}(\Lambda)\big).$$

Exploiting once more the three-term recursion finally shows that the pair (X_0, Λ) satisfies (5.12) and therefore constitutes an invariant pair of the polynomial eigenvalue problem (5.6).

Furthermore, the matrix Y in the invariant pair (Y, Λ) returned by the linear eigensolver has full column rank. In case of the Arnoldi method, the columns of Y are typically even orthonormal. Since we have shown above that $Y = \mathbf{V}_d^\tau(X_0, \Lambda)$, the pair (X_0, Λ) is also minimal by Corollary 3.2.2.

Remark 5.1.1. The preceding discussion suggests a simple extraction procedure: Given an invariant pair (Y, Λ) of the linearization with Y having full column rank, a minimal invariant pair of the polynomial eigenvalue problem is obtained as (X_0, Λ), where X_0 denotes the first block component of Y. In finite-precision arithmetic, this relation is affected by roundoff error. Numerical aspects of such extraction procedures have been discussed in [16] for the class of so-called L_1 linearizations. Additionally, alternative extraction algorithms have been proposed in [16], which turn out to be numerically more robust in certain situations. However, because in the particular setting of this chapter $|\tau_k(\lambda)| \le 1$ for all eigenvalues λ of interest, we expect that X_0 is a dominant component of Y and, hence, a numerically reasonable choice. This is confirmed by our numerical experiments, which also demonstrate that suitable adaptions of the alternative algorithms mentioned above do not result in a significantly improved accuracy.

The accuracy of the extracted minimal invariant pair (X_0, Λ) can be further refined by applying a Newton iteration as described in [16, 83].

5.2 Error analysis

Instead of the original nonlinear eigenvalue problem (1.1), the method developed in Section 5.1 solves the perturbed problem

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

where the perturbation $\triangle T = P - T$ amounts to the interpolation error. It is therefore important to analyze the impact of such a perturbation on the eigenvalues or, more generally, on the minimal invariant pairs. For this purpose, we will derive a general perturbation result for nonlinear eigenvalue problems. This will then be combined with a polynomial approximation result to establish convergence rates for our method.

5.2.1 First-order perturbation theory. In the following, let T_0 be a holomorphic function on some domain $\mathcal{D} \subset \mathbb{C}$ with values in $\mathbb{C}^{n\times n}$. Furthermore, we will assume that T_0 is bounded on $\mathcal{D}$ with respect to the Frobenius norm $\|\cdot\|_{\mathsf{F}}$ and regular; i.e., $\det T_0(\lambda)$ does not vanish identically for all $\lambda \in \mathcal{D}$.

Let $(X_0, \Lambda_0) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ be a minimal invariant pair of T_0 such that all eigenvalues of Λ_0 are contained inside $\mathcal{D}$. Then the triple (X_0, Λ_0, T_0) constitutes a solution of the nonlinear equation

$$F(X, \Lambda, T) = 0 \tag{5.15}$$

with

$$\begin{aligned} &F : \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m} \times B(\mathcal{D}) \to \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}, \\ &(X, \Lambda, T) \mapsto \left(\frac{1}{2\pi \mathrm{i}} \int_\Gamma T(\xi) X (\xi I - \Lambda)^{-1} \,\mathrm{d}\xi, W^{\mathsf{H}} \big[\mathbf{V}_\ell(X, \Lambda) - \mathbf{V}_\ell(X_0, \Lambda_0) \big] \right). \end{aligned} \tag{5.16}$$

Here, Γ is a contour in $\mathcal{D}$ containing the eigenvalues of Λ in its interior, and $B(\mathcal{D})$ denotes the Banach space of all bounded, holomorphic, $\mathbb{C}^{n\times n}$-valued functions on $\mathcal{D} \subset \mathbb{C}$, equipped with the supremum norm

$$\|\cdot\|_\infty : B(\mathcal{D}) \to \mathbb{R}, \quad T \mapsto \|T\|_\infty := \sup_{\lambda\in\mathcal{D}} \|T(\lambda)\|_{\mathsf{F}}.$$

Note that the convergence of functions in this norm amounts to uniform convergence. The first term of $F(X_0, \Lambda_0, T_0) = 0$ as defined in (5.16) characterizes the invariance of the pair (X_0, Λ_0) with respect to the matrix-valued function T_0, whereas the second term characterizes minimality, provided that the normalization matrix $W \in \mathbb{C}^{\ell n\times m}$ is chosen such that $W^{\mathsf{H}} \mathbf{V}_\ell(X_0, \Lambda_0)$ is invertible.

Lemma 5.2.1. *The mapping F defined in* (5.16) *is continuously Fréchet differentiable in a neighborhood of (X_0, Λ_0, T_0).*

Proof. As a norm in the space $\mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m} \times B(\mathcal{D})$, we employ

$$\|(\triangle X, \triangle \Lambda, \triangle T)\| := \|\triangle X\|_{\mathsf{F}} + \|\triangle \Lambda\|_{\mathsf{F}} + \|\triangle T\|_\infty.$$

Since the continuous differentiability of the second component of F is easily seen, we will only treat the first component $F^{(1)}$ and demonstrate that its derivative is given by the map

$$\begin{aligned} \mathrm{D}F^{(1)}(X, \Lambda, T)(\triangle X, \triangle \Lambda, \triangle T) &= \mathrm{D}_X F^{(1)}(X, \Lambda, T)(\triangle X) + \mathrm{D}_\Lambda F^{(1)}(X, \Lambda, T)(\triangle \Lambda) \\ &\quad + \mathrm{D}_T F^{(1)}(X, \Lambda, T)(\triangle T) \end{aligned}$$

with

$$\begin{aligned}
\mathrm{D}_X F^{(1)}(X,\Lambda,T)(\triangle X) &= \frac{1}{2\pi\mathrm{i}}\int_\Gamma T(\xi)\triangle X(\xi I-\Lambda)^{-1}\,\mathrm{d}\xi,\\
\mathrm{D}_\Lambda F^{(1)}(X,\Lambda,T)(\triangle \Lambda) &= \frac{1}{2\pi\mathrm{i}}\int_\Gamma T(\xi)X(\xi I-\Lambda)^{-1}\triangle\Lambda(\xi I-\Lambda)^{-1}\,\mathrm{d}\xi,\\
\mathrm{D}_T F^{(1)}(X,\Lambda,T)(\triangle T) &= \frac{1}{2\pi\mathrm{i}}\int_\Gamma \triangle T(\xi)X(\xi I-\Lambda)^{-1}\,\mathrm{d}\xi.
\end{aligned}$$

For this purpose, let $(X,\Lambda,T)\in\mathbb{C}^{n\times m}\times\mathbb{C}^{m\times m}\times B(\mathcal{D})$ be fixed. We assume Λ to be sufficiently close to Λ_0 so that its eigenvalues still lie inside $\mathcal{D}$. Consequently, there exists a contour Γ in $\mathcal{D}$ which contains all eigenvalues of Λ in its interior. Since the number of eigenvalues is finite, we can, w.l.o.g., assume the contour to possess a finite length L. Let $\gamma:[0,1]\to\Gamma$ be a parameterization of the contour Γ. As Γ touches none of the eigenvalues of Λ, the mapping $\varphi\mapsto\|(\gamma(\varphi)I-\Lambda)^{-1}\|_{\mathsf{F}}$ is continuous and therefore bounded on the compact interval $[0,1]$ by some $M>0$.

Now suppose $\|(\triangle X,\triangle\Lambda,\triangle T)\|<M^{-1}$, particularly implying $\|\triangle\Lambda\|_{\mathsf{F}}<M^{-1}$. Thus, $\|\triangle\Lambda(\xi I-\Lambda)^{-1}\|_{\mathsf{F}}<1$ for any $\xi\in\Gamma$, and the Neumann series gives

$$\begin{aligned}
\bigl[\xi I-(\Lambda+\triangle\Lambda)\bigr]^{-1} &= (\xi I-\Lambda)^{-1}\bigl[I-\triangle\Lambda(\xi I-\Lambda)^{-1}\bigr]^{-1}\\
&= (\xi I-\Lambda)^{-1}\sum_{k=0}^{\infty}\bigl[\triangle\Lambda(\xi I-\Lambda)^{-1}\bigr]^k\\
&= (\xi I-\Lambda)^{-1}+(\xi I-\Lambda)^{-1}\triangle\Lambda(\xi I-\Lambda)^{-1}+O\bigl(\|\triangle\Lambda\|_{\mathsf{F}}^2\bigr),
\end{aligned}$$

where the constant implicitly contained in the $O\bigl(\|\triangle\Lambda\|_{\mathsf{F}}^2\bigr)$ term is independent of ξ. Altogether, we obtain

$$\begin{aligned}
&\bigl\|F^{(1)}(X+\triangle X,\Lambda+\triangle\Lambda,T+\triangle T)-F^{(1)}(X,\Lambda,T)\\
&\qquad -\mathrm{D}F^{(1)}(X,\Lambda,T)(\triangle X,\triangle\Lambda,\triangle T)\bigr\|_{\mathsf{F}}\\
&=\left\|\frac{1}{2\pi\mathrm{i}}\int_\Gamma[\triangle T(\xi)X+T(\xi)\triangle X+\triangle T(\xi)\triangle X](\xi I-\Lambda)^{-1}\triangle\Lambda(\xi I-\Lambda)^{-1}\right.\\
&\qquad\left.+\triangle T(\xi)\triangle X(\xi I-\Lambda)^{-1}+O\bigl(\|\triangle\Lambda\|_{\mathsf{F}}^2\bigr)\,\mathrm{d}\xi\right\|_{\mathsf{F}}\\
&=O\bigl(\|(\triangle X,\triangle\Lambda,\triangle T)\|^2\bigr)
\end{aligned}$$

confirming the claim that $F^{(1)}$ is differentiable with the derivative $\mathrm{D}F^{(1)}$ stated above. The continuity of $\mathrm{D}F^{(1)}$ can be established by a similar estimate. □

We now consider the derivative of F at (X_0,Λ_0,T_0) only with respect to X and Λ, but not T. The corresponding linear operator from $\mathbb{C}^{n\times m}\times\mathbb{C}^{m\times m}$ onto itself will be denoted by $\mathrm{D}_{(X,\Lambda)}F(X_0,\Lambda_0,T_0)$. The next result is an adaption of Theorem 4.1.2 to the setting of this chapter, where F additionally depends on T. Its proof using Theorem 4.1.2 is straightforward and therefore omitted.

Theorem 5.2.2 ([83, Theorem 10])**.** *Let (X_0,Λ_0) be a minimal invariant pair of T_0. Then the derivative $\mathrm{D}_{(X,\Lambda)}F$ at (X_0,Λ_0,T_0) is an automorphism on $\mathbb{C}^{n\times m}\times\mathbb{C}^{m\times m}$ if and only if (X_0,Λ_0) is simple.*

Combining Lemma 5.2.1 and Theorem 5.2.2 permits us to apply the Implicit Function Theorem to the nonlinear equation (5.15) in the vicinity of a simple invariant pair (X_0, Λ_0) of T_0. This yields the existence of continuously differentiable functions $X : B(\mathcal{D}) \to \mathbb{C}^{n\times m}$ and $\Lambda : B(\mathcal{D}) \to \mathbb{C}^{m\times m}$ with $X(T_0) = X_0$ and $\Lambda(T_0) = \Lambda_0$ such that

$$F\big(X(T), \Lambda(T), T\big) = 0$$

for all T in a neighborhood of T_0. Moreover, the derivatives with respect to T of these two functions at T_0 are given by

$$\begin{bmatrix} \mathrm{D}_T X(T_0) \triangle T_0 \\ \mathrm{D}_T \Lambda(T_0) \triangle T_0 \end{bmatrix} = -\Big(\big[\mathrm{D}_{(X,\Lambda)} F(X_0, \Lambda_0, T_0)\big]^{-1} \circ \mathrm{D}_T F(X_0, \Lambda_0, T_0) \Big) \triangle T_0,$$

where $\big[\mathrm{D}_{(X,\Lambda)} F(X_0, \Lambda_0, T_0)\big]^{-1}$ refers to the inverse of the bijective linear operator $\mathrm{D}_{(X,\Lambda)} F(X_0, \Lambda_0, T_0) : \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m} \to \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ as established by Theorem 5.2.2. Setting $T = T_0 + \triangle T_0$, we conclude that the perturbed nonlinear eigenvalue problem $(T_0 + \triangle T_0)(\lambda)x = 0$ possesses a minimal invariant pair (X, Λ) satisfying

$$\begin{bmatrix} X \\ \Lambda \end{bmatrix} = \begin{bmatrix} X_0 \\ \Lambda_0 \end{bmatrix} - \Big(\big[\mathrm{D}_{(X,\Lambda)} F(X_0, \Lambda_0, T_0)\big]^{-1} \circ \mathrm{D}_T F(X_0, \Lambda_0, T_0) \Big) \triangle T_0 + o\big(\|\triangle T_0\|_\infty\big). \tag{5.17}$$

The main result of this section is summarized in the subsequent theorem.

Theorem 5.2.3. *Let T_0 and $\triangle T_0$ be bounded, holomorphic, $\mathbb{C}^{n\times n}$-valued functions on some domain $\mathcal{D} \subset \mathbb{C}$ and suppose that T_0 is regular. Let $(X_0, \Lambda_0) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ be a simple invariant pair of T_0. If $\|\triangle T_0\|_\infty$ is sufficiently small, then there exists a minimal invariant pair $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ of the perturbed function $T_0 + \triangle T_0$ satisfying* (5.17) *with F defined as in* (5.16).

Remark 5.2.4. Theorem 5.2.3 suggests that the norm of the linear operator

$$\big[\mathrm{D}_{(X,\Lambda)} F(X_0, \Lambda_0, T_0)\big]^{-1} \circ \mathrm{D}_T F(X_0, \Lambda_0, T_0)$$

can be regarded as a condition number for the simple invariant pair (X_0, Λ_0).

5.2.2 Convergence rates. We will now apply Theorem 5.2.3 to analyze the method from Section 5.1. To this end, we assume that T is a regular, analytic function on the real interval $[-1, 1]$ with values in $\mathbb{C}^{n\times n}$ and, thus, can be extended to a holomorphic function on a neighborhood of this interval in the complex plane. For simplicity, this extension will also be referred to as T. In particular, we can choose $\rho > 1$ such that the Bernstein ellipse

$$E_\rho := \big\{ \cos(t - \mathrm{i} \ln \bar{\rho}) : t \in [0, 2\pi],\ \bar{\rho} \in [1, \rho] \big\}$$

is contained in the analyticity domain of T. Moreover, the holomorphic extension of T obviously inherits its regularity. With this notation, we obtain the following convergence result.

Corollary 5.2.5. *Let T be as above and let $P^{(d)}$ denote the interpolating polynomial of degree d for T with respect to the Chebyshev nodes of either the first or the second kind. Let (X, Λ) be a simple invariant pair of T such that all eigenvalues of Λ lie in the real interval $[-1, 1]$. Then there exists a sequence of minimal invariant pairs (X_d, Λ_d) belonging to the polynomials $P^{(d)}$ which converges to (X, Λ) exponentially as $d \to \infty$.*

Proof. Choose $\mathcal{D} = E_{\rho_0}$ with $1 < \rho_0 < \rho$ as above, and set $\triangle T^{(d)} := P^{(d)} - T$ for all $d \in \mathbb{N}_0$. Because $\mathcal{D}$ is compact, we have $T \in B(\mathcal{D})$ as well as $\triangle T^{(d)} \in B(\mathcal{D})$ for all $d \in \mathbb{N}_0$. Consequently, by Theorem 5.2.3, there exists a minimal invariant pair (X_d, Λ_d) of $P^{(d)}$ satisfying

$$\left\| \begin{bmatrix} X_d \\ \Lambda_d \end{bmatrix} - \begin{bmatrix} X \\ \Lambda \end{bmatrix} \right\|_{\mathsf{F}} = O\big(\|\triangle T^{(d)}\|_{\mathsf{F}}\big).$$

Since the interpolation error $\triangle T^{(d)}$ converges to zero exponentially according to Proposition 5.2.6 below, the assertion follows. □

The proof of Corollary 5.2.5 relies on convergence estimates for the Chebyshev interpolant inside the Bernstein ellipse E_{ρ_0}. These will be covered by the subsequent proposition, which is a variation of classical polynomial approximation results; see, e.g., [90].

Proposition 5.2.6. *Let $T : U \to \mathbb{C}^{n\times m}$ be holomorphic in a neighborhood U of the Bernstein ellipse E_ρ with $\rho > \rho_0 > 1$ and let $P^{(d)}$ denote the interpolating polynomial of degree d for T with respect to the Chebyshev nodes of either the first or the second kind. Then there exists a constant $C > 0$ depending only on T, ρ, and ρ_0 such that for all $\lambda \in E_{\rho_0}$,*

$$\|T(\lambda) - P^{(d)}(\lambda)\|_{\mathsf{F}} \le C\Big(\frac{\rho_0}{\rho}\Big)^d.$$

Proof. Depending on what kind of Chebyshev nodes are used, we define $(Q_d)_{d\in\mathbb{N}_0}$ to be the sequence of Chebyshev polynomials of either the first or the second kind. In any event, the interpolation nodes are the zeroes of the polynomials Q_d.

In the following, we will show the claim for the Chebyshev nodes of the first kind. In this case, it is well known that

$$Q_d(\cos\theta) = \cos(d\theta). \tag{5.18}$$

The statement for the Chebyshev nodes of the second kind then follows by similar arguments using the identity

$$Q_d(\cos\theta)\sin\theta = \sin\big((d+1)\theta\big)$$

instead and is therefore omitted.

Let $\lambda \in E_{\rho_0}$. If λ is identical with one of the interpolation nodes, the claimed inequality trivially holds true for any $C > 0$. Hence, we may assume, w.l.o.g., that $Q_d(\lambda) \neq 0$. By applying the residue theorem to the function $\xi \mapsto \frac{T(\xi)}{(\xi-\lambda)Q_d(\xi)}$ and exploiting that all roots of its denominator are simple, one shows that the interpolation error satisfies

$$T(\lambda) - P^{(d)}(\lambda) = \frac{Q_d(\lambda)}{2\pi\mathrm{i}} \int_{\partial E_\rho} \frac{T(\xi)}{(\xi-\lambda)Q_d(\xi)}\,\mathrm{d}\xi, \tag{5.19}$$

which is a matrix version of Hermite's theorem [33, Theorem 3.6.1].

We proceed by estimating the individual factors in the contour integral representation of the interpolation error on the right-hand side of Equation (5.19). To begin with, we notice that $\xi \in \partial E_\rho$ can be expressed as $\xi = \cos(t - \mathrm{i}\ln\rho)$ for some $t \in [0, 2\pi]$, and hence $Q_d(\xi) = \cos(dt - \mathrm{i}d\ln\rho)$ due to (5.18). A simple calculation then reveals that

$$|Q_d(\xi)|^2 = \tfrac{1}{4}(\rho^d - \rho^{-d})^2 + \cos^2(dt),$$

implying the estimate

$$\tfrac{1}{2}(\rho^d - \rho^{-d}) \le |Q_d(\xi)| \le \tfrac{1}{2}(\rho^d + \rho^{-d})$$

because dt is real. Analogously, $\lambda \in E_{\rho_0}$ can be written as $\lambda = \cos(s - \mathrm{i}\ln\bar{\rho})$ for some $s \in [0, 2\pi]$, $\bar{\rho} \in [1, \rho_0]$, and we conclude that

$$|Q_d(\lambda)| \le \tfrac{1}{2}(\bar{\rho}^d + \bar{\rho}^{-d}) \le \tfrac{1}{2}(\rho_0^d + \rho_0^{-d}).$$

Furthermore, $|\xi - \lambda|$ is bounded from below by the minimal distance between E_{ρ_0} and ∂E_ρ, which is given by $\operatorname{dist}(E_{\rho_0}, \partial E_\rho) = \frac{1}{2}[\rho + \rho^{-1} - (\rho_0 + \rho_0^{-1})]$ due to geometric considerations. Finally, $\|T(\xi)\|_{\mathsf{F}} \le \|T\|_\infty$. Taking Frobenius norms in (5.19) and inserting the above estimates, we obtain the bound

$$\big\|T(\lambda) - P^{(d)}(\lambda)\big\|_{\mathsf{F}} \le \frac{L_\rho \|T\|_\infty}{2\pi \operatorname{dist}(E_{\rho_0}, \partial E_\rho)} \cdot \frac{\rho_0^d + \rho_0^{-d}}{\rho^d - \rho^{-d}},$$

where L_ρ is the circumference of the Bernstein ellipse E_ρ. The proof is completed by taking into account that $\rho_0^{-d} \to 0$ and $\rho^{-d} \to 0$ as $d \to \infty$. □

5.2.3 Spurious eigenvalues.

Based on real interpolation nodes, the interpolating polynomials $P^{(d)}$ tend to be accurate only in the vicinity of the real axis. Away from the real axis, the approximation quality quickly deteriorates. This might cause the appearance of spurious eigenvalues, i.e., eigenvalues of the interpolating polynomial which do not approximate any eigenvalue of the original nonlinear eigenvalue problem in the sense that their associated residual is large. However, the subsequent result shows that this problem does not occur for sufficiently large polynomial degree d.

Corollary 5.2.7. *Let the assumptions of Proposition 5.2.6 hold and let $\lambda \in E_{\rho_0}$ be such that $T(\lambda)$ is nonsingular (i.e., λ is not an eigenvalue of T). Then, there exists a nonnegative integer d_0 such that $P^{(d)}(\lambda)$ is nonsingular (i.e., λ is not an eigenvalue of $P^{(d)}$) for all $d \ge d_0$.*

Proof. Let $\lambda \in E_{\rho_0}$ be fixed. According to Proposition 5.2.6, we can choose $d_0 \in \mathbb{N}_0$ such that the Frobenius norm of the interpolation error $\triangle T^{(d)}(\lambda) = P^{(d)}(\lambda) - T(\lambda)$ is strictly bounded from above by $\|T(\lambda)^{-1}\|_{\mathsf{F}}^{-1}$ for all $d \ge d_0$. Consequently, we have $\|-\triangle T^{(d)}(\lambda) T(\lambda)^{-1}\|_{\mathsf{F}} < 1$ for all $d \ge d_0$, and a Neumann series argument therefore demonstrates that $P^{(d)}(\lambda) = T(\lambda) + \triangle T^{(d)}(\lambda)$ is invertible with

$$P^{(d)}(\lambda)^{-1} = T(\lambda)^{-1} \sum_{k=0}^{\infty} \big[-\triangle T^{(d)}(\lambda) T(\lambda)^{-1}\big]^k. \qquad \square$$

Corollary 5.2.7 states that in the limit $d \to \infty$, $\lambda \in E_{\rho_0}$ can only be an eigenvalue of $P^{(d)}$ if it is also an eigenvalue of T. Thus, asymptotically, there will be no spurious eigenvalues inside the Bernstein ellipse E_{ρ_0}. Since the interval $[-1, 1]$ is enclosed by E_{ρ_0}, we expect spurious eigenvalues to occur only in some distance

no.	eigenvalue	multiplicity
1	5.441398	1
2	7.695299	3
3	9.424778	3
4	10.419484	3
5	10.882796	1
6	11.754763	6

(a)

no.	eigenvalue
1	6.484702318577543
2	8.142495692472265
3	8.142499335034771
4	9.053846829423080
5	9.716892649192921
6	9.716894006586880

(b)

Table 5.1: (a) The 6 smallest eigenvalues of the negative Laplace operator on the unit cube. (b) Reference eigenvalues for the Fichera corner computed by the method from Section 5.1 using an interpolating polynomial of degree 30 on a uniform boundary mesh with 2400 triangles.

h	no. of triangles	average execution time (s)
$1/6$	864	409
$1/8$	1536	1154
$1/10$	2400	2722

Table 5.2: Execution times for determining all eigenvalues within the interval of interest using an interpolating polynomial of degree 30 for increasingly fine boundary meshes of the unit cube.

to the interval. This motivates the following mechanism for detecting spurious eigenvalues: An eigenvalue is discarded as spurious if its real part lies outside the interval $[-1, 1]$ or its imaginary part exceeds a certain threshold in magnitude. In particular for nonlinear eigenvalue problems resulting from boundary integral formulations (see Section 2.3), this strategy leads to considerable computational savings over the conventional approach of checking the residuals for all computed eigenvalues.

5.3 Numerical Experiments

To assess the performance of the method developed in Section 5.1, we have applied it to a set of test problems. All computations have been performed under MATLAB 7.10 (R2010a) on a cluster of 24 Intel Xeon X5650 processors with 72 GB of shared memory. The reported computing times are averages over 20 identical runs. For the solution of the linearized eigenvalue problems, we have utilized a MATLAB implementation of the Arnoldi algorithm with Krylov-Schur restarting [125].

5.3.1 Unit cube. In our first experiment, we consider the Laplace eigenvalue problem (2.9) on the unit cube $\Omega = [0, 1]^3$ with homogeneous Dirichlet boundary conditions. The eigenvalues and eigenfunctions of this problem are known to be given by

$$\lambda_{j_1,j_2,j_3} = \pi\sqrt{j_1^2 + j_2^2 + j_3^2},$$
$$u_{j_1,j_2,j_3}(x_1, x_2, x_3) = \sin(j_1\pi x_1)\sin(j_2\pi x_2)\sin(j_3\pi x_3)$$

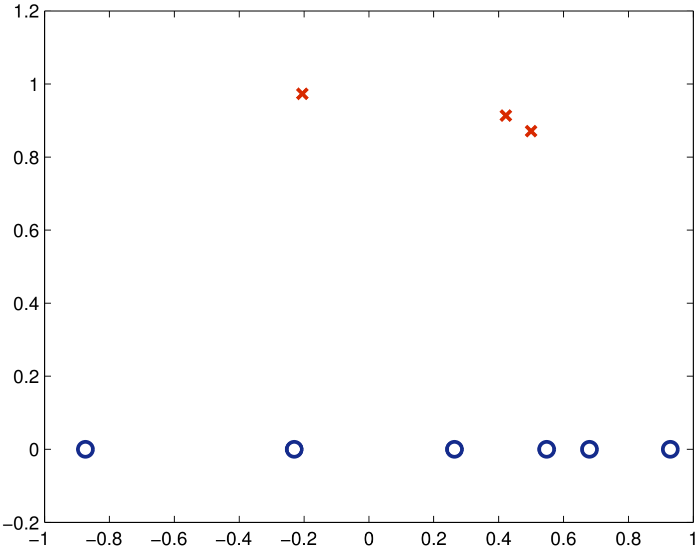

Figure 5.1: Ritz values obtained by the method in Section 5.1 for the Laplace eigenproblem on the unit cube using a uniform boundary mesh with 2400 triangles. The circles mark Ritz values corresponding to true eigenvalues, whereas the crosses indicate spurious eigenvalues.

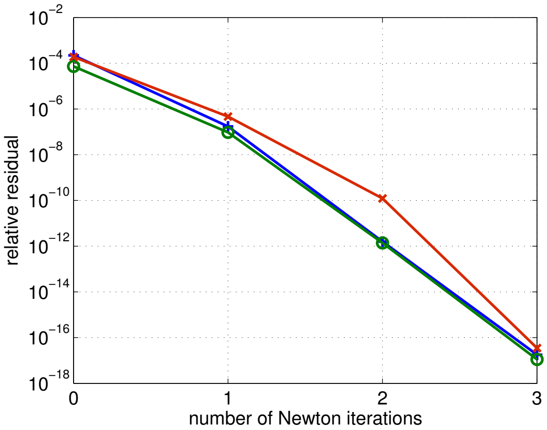

Figure 5.2: Relative residual of an invariant pair representing the first 11 eigenvalues of the Laplace eigenvalue problem on the unit cube during three steps of Newton-based iterative refinement. Each color represents a different level of mesh refinement: $h = \frac{1}{6}$, 864 triangles (blue), $h = \frac{1}{8}$, 1536 triangles (green), $h = \frac{1}{10}$, 2400 triangles (red).

h	no. of triangles	average execution time (s)
$1/6$	864	361
$1/8$	1536	1054
$1/10$	2400	2451

Table 5.3: Execution times for determining all eigenvalues within the interval of interest using an interpolating polynomial of degree 30 for increasingly fine boundary meshes of the Fichera corner.

for $j_1, j_2, j_3 = 1, 2, \ldots$. The 6 smallest eigenvalues are summarized in Table 5.1(a). The occurence of multiple eigenvalues is due to the symmetry of the domain.

We construct a boundary formulation of the problem as described in Section 2.3 and solve the resulting nonlinear eigenvalue problem by the method developed in Section 5.1. To capture the six smallest distinct eigenvalues (17, when counting multiplicities), we select $[5, 12]$ as the interval of interest. Furthermore, we set the degree of the interpolating polynomial to 12 and compute 20 Ritz values with the implicitly restarted Arnoldi algorithm. The result for a uniform boundary mesh with 2400 triangles is depicted in Figure 5.1. The plot also reveals a small number of spurious eigenvalues (marked by crosses). However, as predicted by the results in Section 5.2.3, these spurious eigenvalues are well-separated from the true eigenvalues close to the real axis and can be easily identified.

We have experimented with different levels of mesh refinement and different degrees of the interpolating polynomial. Figure 5.4 shows the spectral convergence of the computed eigenvalues towards a reference solution obtained with polynomial degree 30. The numerical results support the exponential convergence of eigenvalues predicted by Corollary 5.2.5. Also note that all eigenvalues converge at roughly the same rate. The execution times for the reference solutions are reported in Table 5.2. These include the time for setting up the interpolating polynomial, which constitutes the dominating part, as well as the time for the solution of the polynomial eigenvalue problem.

Furthermore, we have implemented and tested the extraction scheme for minimal invariant pairs as well as their subsequent refinement via Newton iterations outlined in Remark 5.1.1. For different levels of mesh refinement, we apply the Arnoldi method to compute an approximate minimal invariant pair representing the first 11 eigenvalues of a degree-20 interpolating polynomial. These initial minimal invariant pairs have relative residuals of about 10^{-4}. We then perform three Newton steps. The result is depicted in Figure 5.2. Already after two steps, the relative residual has decreased to an order between 10^{-10} and 10^{-12}. Finally, after the third step, the residual reaches machine accuracy.

5.3.2 Fichera corner. As a second experiment, we consider the Laplace eigenvalue problem in (2.9) with homogeneous Dirichlet boundary conditions for the Fichera corner $\Omega = [0, 1]^3 \setminus [1/2, 1]^3$. The boundary element formulation as well as the resulting nonlinear eigenvalue problem for this case are again obtained as outlined in Section 2.3. However, this time there is no analytic expression for the eigenvalues available.

On a uniform boundary mesh with 2400 triangles and with an interpolation polynomial of degree 30, our new method computes the approximate eigenvalues listed in Table 5.1(b). The spectral convergence of these eigenvalues towards a

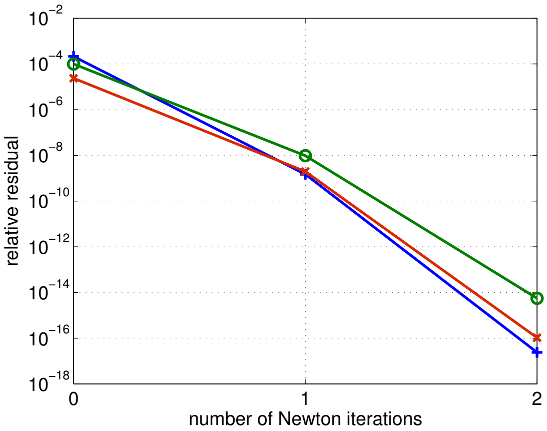

Figure 5.3: Relative residual of an invariant pair representing the first 6 eigenvalues of the Laplace eigenvalue problem for the Fichera corner during two steps of Newton-based iterative refinement. Each color represents a different level of mesh refinement: $h = \frac{1}{6}$, 864 triangles (blue), $h = \frac{1}{8}$, 1536 triangles (green), $h = \frac{1}{10}$, 2400 triangles (red).

reference solution computed with polynomial degree 30 is illustrated in Figure 5.5. Once more, exponential convergence of the eigenvalues is observed, in agreement with the statement of Corollary 5.2.5. Table 5.3 summarizes the computing times for the reference solutions, consisting of the time spent to set up the interpolating polynomial and the time for solving the polynomial eigenvalue problem.

Also in this case, we have applied the extraction scheme and Newton-based iterative refinement from Remark 5.1.1. Starting from an approximate minimal invariant pair with relative residual 10^{-4} of a degree-20 interpolating polynomial, the first refinement step brings the residual down to about 10^{-9}. Already after the second step, the relative residual approaches the level of the machine accuracy. The results are visualized in Figure 5.3.

Remark 5.3.1. When considering the execution times in Tables 5.2 and 5.3, one should take into account that we did not use a highly optimized BEM code for our computations. Possible improvements include, e.g., the exploitation of the inherent parallelism in the computation of the matrix entries (2.11) as well as the use of hierarchical matrix techniques; see, e.g., [13, 54].

Contributions within this chapter

In this chapter, we have developed a solution technique for general holomorphic eigenvalue problems using polynomial approximation of the matrix-valued function. The resulting polynomial eigenvalue problem is then solved by applying a Krylov subspace eigensolver to a suitable linearization. This approach is especially advantageous in situations where evaluating the matrix-valued function is very costly because the number of such operations is kept to a minimum. In lieu of Taylor expansion as often encountered in applications, we approximate the matrix-valued function through polynomial interpolation in Chebyshev nodes. This leads

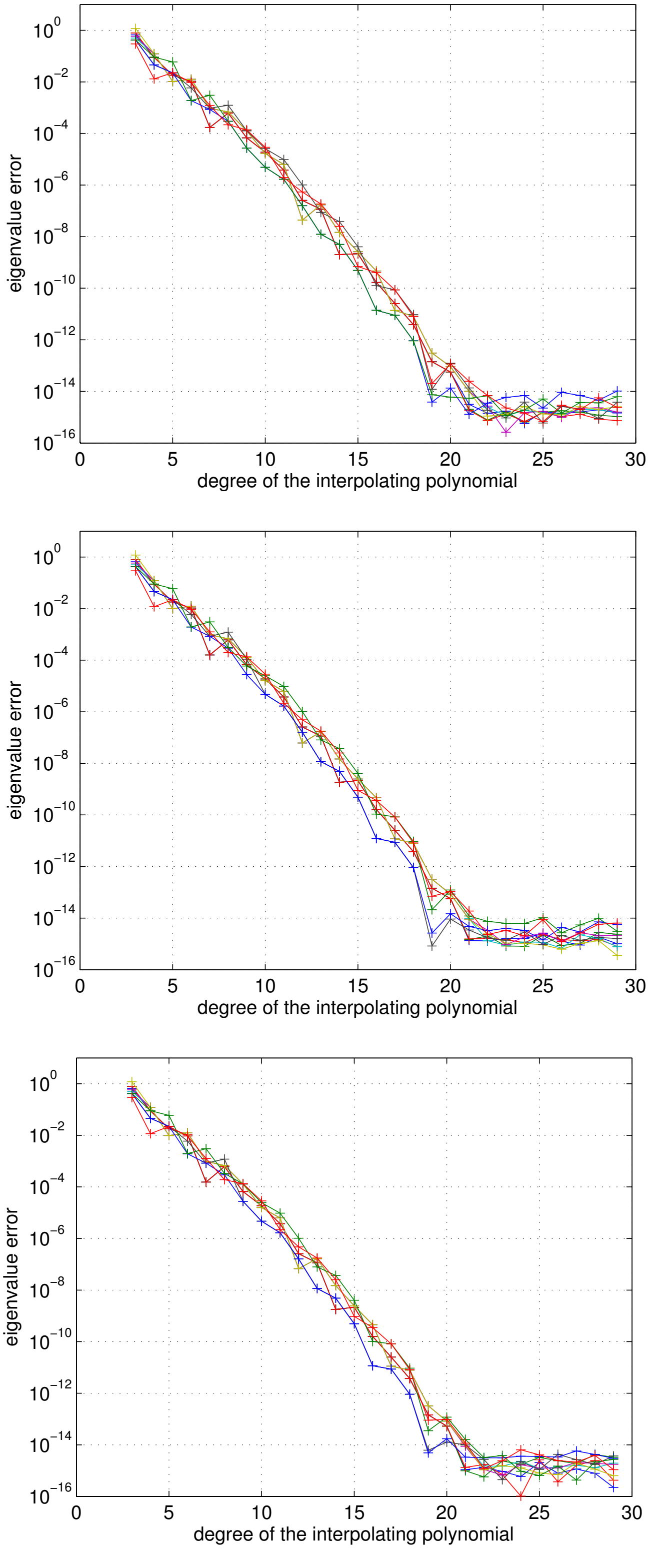

Figure 5.4: Spectral convergence of the first 10 eigenvalues for increasingly fine boundary meshes of the unit cube. Top. $h = \frac{1}{6}$, 864 triangles. Center. $h = \frac{1}{8}$, 1536 triangles. Bottom. $h = \frac{1}{10}$, 2400 triangles.

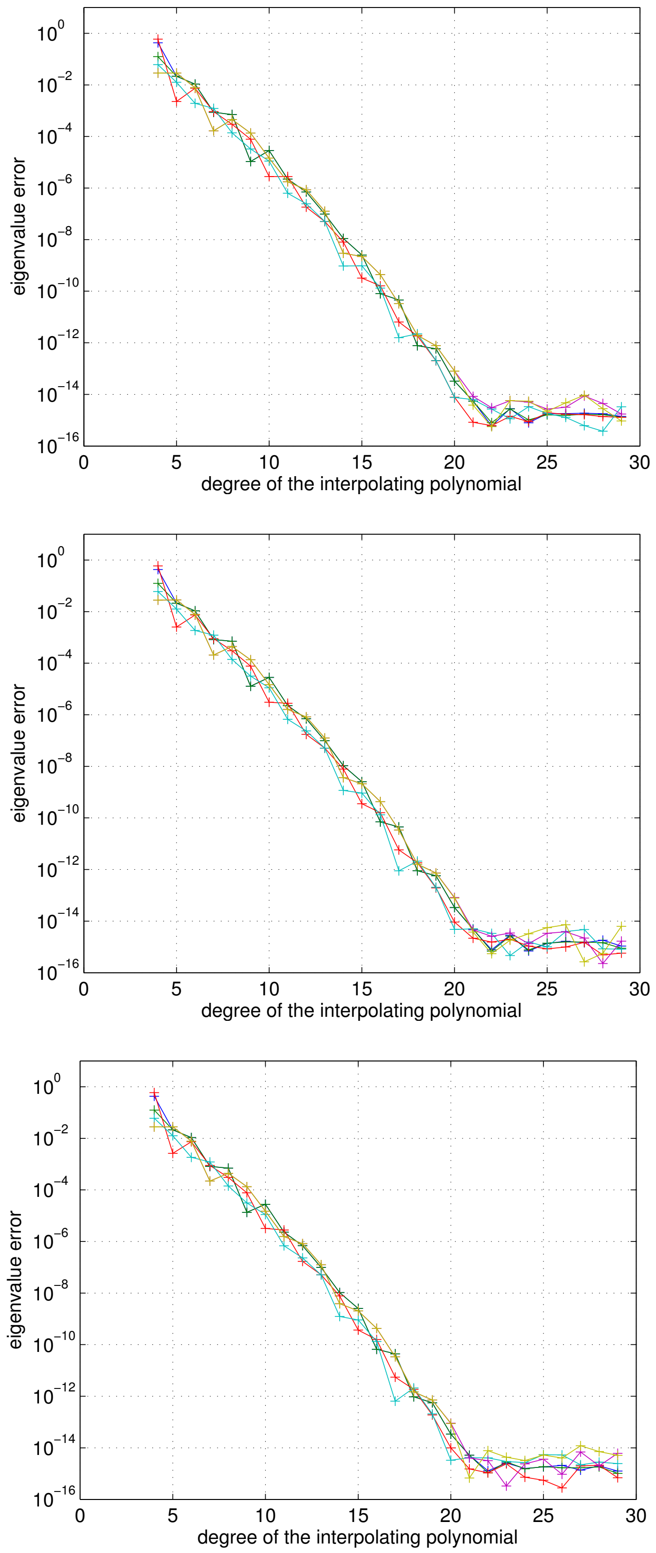

Figure 5.5: Spectral convergence of the first 6 eigenvalues for increasingly fine boundary meshes of the Fichera corner. Top. $h = \frac{1}{6}$, 864 triangles. Center. $h = \frac{1}{8}$, 1536 triangles. Bottom. $h = \frac{1}{10}$, 2400 triangles.

to exponential convergence of the computed eigenvalues with respect to the number of interpolation nodes if the eigenvalues of interest lie on or close to the real axis or any other predetermined curve in the complex plane.

In Section 5.1, we derive the method. In particular, in Section 5.1.3, we discuss the efficient solution of the linear system arising in the Krylov subspace eigensolver applied to the linearized polynomial eigenvalue problem. In Section 5.1.4, we comment on the extraction of minimal invariant pairs for the polynomial eigenvalue problem from minimal invariant pairs for the linearization.

In Section 5.2, we analyze the error incurred by the polynomial interpolation. To this end, we derive a first-order perturbation expansion for simple invariant pairs of a general holomorphic eigenvalue problem in Theorem 5.2.3. Combining this expansion with standard convergence results for Chebyshev interpolation, we show in Corollary 5.2.5 that the error in the computed minimal invariant pairs decays exponentially with the number of Chebyshev nodes. Finally, we address the issue of spurious eigenvalues in Section 5.2.3. Corollary 5.2.7 demonstrates that spurious eigenvalue can only occur pre-asymptotically and eventually vanish in the asymptotic regime. All of the above-mentioned results have been published by Kressner and the author in [40].

Chapter 6

Deflation techniques for nonlinear eigenvalue problems

Unfortunately, most available algorithms for the solution of the nonlinear eigenvalue problem (1.1) are single-vector iterations and therefore directed towards computing one eigenpair only. The robust and reliable calculation of several eigenpairs—although required by many applications (see Chapter 2)—is much more difficult and less well studied.

Essentially, there appear to be two different strategies for the computation of several eigenpairs: Either they are computed simultaneously or successively. The simultaneous computation of eigenpairs can be achieved by deriving block versions of the aforementioned single-vector iterations. An example is the block Newton method constructed in [83], which utilizes minimal invariant pairs (see Chapter 3) as a numerically robust means of representing several eigenpairs. However, the resulting methods are harder to analyze and seem to be more restrictive in terms of local convergence than their single-vector counterparts. Moreover, the number of eigenpairs to be approximated must be known in advance.

In contrast, computing eigenpairs successively avoids all of the above disadvantages. In particular, convergence can be monitored and steered much more easily for individual eigenpairs than for blocks. However, one has to ensure that the algorithm does not repeatedly converge to the same eigenpair. In principle, this issue could be addressed by keeping a list of previously converged eigenpairs and then discarding further copies as they arise but such a strategy seems impractical for several reasons. First of all, it does not save the computational work spent on recomputing eigenpairs. More importantly, it is likely to cause difficulties in the presence of multiple or nearly-multiple eigenvalues.

A much more preferable solution would be to deflate converged eigenpairs from the problem as this reliably prevents reconvergence and tends to enlarge the convergence basin for unconverged eigenpairs. This is a well-known technique for linear eigenvalue problems [46, 86, 125, 10], where the fact that their eigenvectors are linearly independent is exploited to accomplish the deflation via partial Schur forms. For nonlinear eigenvalue problems, though, linear independence of the eigenvectors is no longer guaranteed. Insisting that the computed eigenvectors be linearly independent therefore bears the danger of missing eigenpairs, making deflation a much more delicate task in this context.

Most existing deflation approaches for nonlinear eigenvalue problems are based on a linear reformulation of the problem, which can then be deflated by established techniques. Most notably, such linearizations are known and widely used for polynomial eigenvalue problems; see, e.g., [48, 93]. In this case it has been shown [95] that the deflation can be accomplished in an implicit fashion, sparing the need to actually set up the (often considerably larger) linearization. Recently, a method has been proposed [67] in connection with the infinite Arnoldi algorithm (see Section 1.2.3) which also enables the reformulation of eigenvalue problems with more general nonlinearities, leading to linear operator eigenvalue problems on an infinite-dimensional vector space. Although this method seems to be applicable to a broad class of nonlinear eigenvalue problems, it requires the problem to be given in an analytical form, which may not be readily available.

In contrast, the nonequivalence deflation developed in [52] for quadratic eigenvalue problems and extended to the general nonlinear case in [51] operates directly on the nonlinear formulation of the problem. The method proceeds by modifying the eigenvalue problem in order to relocate the computed eigenvalues to infinity. Whereas the remaining eigenvalues are not changed, their corresponding eigenvectors are, necessitating additional linear system solves for eigenvector recovery. The applicability of the method appears to be limited to semi-simple eigenvalues. Moreover, the method can be expected to suffer from numerical instabilities in the presence of clustered eigenvalues. The authors demonstrate how to work around these numerical difficulties for polynomial eigenvalue problems only.

In this chapter, we will propose a new deflation strategy for general nonlinear eigenvalue problems, which avoids many of the aforementioned disadvantages. In particular, our approach operates directly on the original nonlinear problem, can handle any type of (holomorphic) nonlinearity, and does not require access to an analytic formulation. The algorithm to be developed computes several eigenpairs successively but represents them together as a single invariant pair. By maintaining the minimality of this invariant pair, we prevent the algorithm from reconverging to the same eigenpairs while retaining the favorable convergence properties of single-vector iterations. Multiple or even defective eigenvalues, on the other hand, are detected and the correct number of copies is retrieved.

6.1 Robust expansion of minimal invariant pairs

Suppose that a minimal invariant pair $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ of the nonlinear eigenvalue problem (1.1) is known. Proposition 3.4.3 demonstrates that if (X, Λ) is non-simple, we can extend it into a minimal invariant pair

$$(\hat{X}, \hat{\Lambda}) = \left(\begin{bmatrix} X & y \end{bmatrix}, \begin{bmatrix} \Lambda & z \\ 0 & \theta \end{bmatrix} \right) \in \mathbb{C}^{n\times(m+1)} \times \mathbb{C}^{(m+1)\times(m+1)} \tag{6.1}$$

of larger size. However, it is easy to see using Proposition 3.4.2 that the ansatz (6.1) is also reasonable under more general circumstances. Hence, our goal becomes the determination of $y \in \mathbb{C}^n$, $z \in \mathbb{C}^m$, and $\theta \in \mathcal{D}$ such that the pair $(\hat{X}, \hat{\Lambda})$ is both invariant and minimal. This strategy can be seen as the nonlinear counterpart to expanding a partial Schur decomposition, such as in the Jacobi-Davidson QR and QZ methods [46].

Throughout this section, Γ denotes a contour in $\mathcal{D}$, enclosing both θ and the eigenvalues of Λ in its interior. The following lemma provides a necessary and sufficient criterion for the invariance of the extended pair $(\hat{X}, \hat{\Lambda})$.

Lemma 6.1.1. *Let (X, Λ) be an invariant pair of the nonlinear eigenvalue problem* (1.1)*. Then the extended pair $(\hat{X}, \hat{\Lambda})$ defined in* (6.1) *is invariant if and only if*

$$T(\theta)y + U(\theta)z = 0, \tag{6.2}$$

where

$$U(\theta) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma T(\xi) X (\xi I - \Lambda)^{-1} (\xi - \theta)^{-1} \,\mathrm{d}\xi. \tag{6.3}$$

Proof. Recall from Definition 3.1.1 that the extended pair $(\hat{X}, \hat{\Lambda})$ is invariant if and only if $\mathbf{T}(\hat{X}, \hat{\Lambda}) = 0$. By Proposition 3.3.7, this condition decomposes into $\mathbf{T}(X, \Lambda) = 0$ as well as $\mathbf{T}(y, \theta) + \mathrm{D}_{[\Lambda,\theta]}\mathbf{T}(X, \cdot)z = 0$. Since (X, Λ) constitutes an invariant pair, the former equation will always be satisfied. Hence, the invariance of $(\hat{X}, \hat{\Lambda})$ is equivalent to the latter equation. The proof is concluded by noticing that the latter equation coincides with (6.2) because $\mathbf{T}(y, \theta) = T(\theta)y$ by definition and $\mathrm{D}_{[\Lambda,\theta]}\mathbf{T}(X, \cdot)z = U(\theta)z$ with $U(\theta)$ as defined in (6.3) by Lemma 3.3.5. □

Note that the left-hand side of Equation (6.2) is linear in both y and z; the dependence on θ is nonlinear but holomorphic as the subsequent result shows.

Lemma 6.1.2. *$U(\theta)$ as defined in* (6.3) *depends holomorphically on θ.*

Proof. The differentiability of $U(\theta)$ is evident from its contour integral representation (6.3). The k-th derivative is given by

$$U^{(k)}(\theta) = \frac{k!}{2\pi \mathrm{i}} \int_\Gamma T(\xi) X (\xi I - \Lambda)^{-1} (\xi - \theta)^{-(k+1)} \,\mathrm{d}\xi. \qquad \square$$

Suppose we have found $(y, z, \theta) \in \mathbb{C}^n \times \mathbb{C}^m \times \mathcal{D}$ such that condition (6.2) is met. We will first give a preparatory result and then state a necessary and sufficient condition for the minimality of the ensuing augmented pair $(\hat{X}, \hat{\Lambda})$.

Lemma 6.1.3. *Let $(\hat{X}, \hat{\Lambda})$ be defined as in* (6.1) *with $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$. For some positive integer ℓ, let the polynomials $p_0, \dots, p_\ell$, given by*

$$p_i(\lambda) = \alpha_i \cdot (\lambda - \beta_{i,1}) \cdots (\lambda - \beta_{i,d_i}), \qquad i = 0, \dots, \ell,$$

constitute a basis for the vector space of polynomials of degree at most ℓ, and define the polynomials $q_0, \dots, q_\ell$ by

$$q_i(\lambda) = \alpha_i \cdot \sum_{j=1}^{d_i} (\Lambda - \beta_{i,1} I) \cdots (\Lambda - \beta_{i,j-1} I) \cdot (\lambda - \beta_{i,j+1}) \cdots (\lambda - \beta_{i,d_i}), \quad i = 0, \dots, \ell.$$

Then,

$$\mathbf{V}^p_{\ell+1}(\hat{X}, \hat{\Lambda}) = \big[\mathbf{V}^p_{\ell+1}(X, \Lambda),\ \mathbf{v}^p(y, z, \theta)\big],$$

where the matrices $\mathbf{V}^p_{\ell+1}(\hat{X}, \hat{\Lambda})$ and $\mathbf{V}^p_{\ell+1}(X, \Lambda)$ are defined as in (3.10) *and*

$$\mathbf{v}^p(y, z, \theta) = \begin{bmatrix} p_0(\theta)y + Xq_0(\theta)z \\ \vdots \\ p_\ell(\theta)y + Xq_\ell(\theta)z \end{bmatrix}. \tag{6.4}$$

Proof. By induction on the polynomial degree d_i of p_i one demonstrates that

$$p_i\left(\begin{bmatrix}\Lambda & z\\ 0 & \theta\end{bmatrix}\right) = \begin{bmatrix}p_i(\Lambda) & q_i(\theta)z\\ 0 & p_i(\theta)\end{bmatrix}$$

for all $i = 0, \dots, \ell$. From this, we conclude $(\mathrm{D}_{[\Lambda,\,\theta]}p_i)z = q_i(\theta)z$ for all $i = 0, \dots, \ell$ by virtue of Proposition 3.3.4. The claim now follows from the latter with the aid of Lemma 3.3.6 and Proposition 3.3.7. □

Lemma 6.1.4. *Let the assumptions of Lemma 6.1.3 be fulfilled and assume that $(X, \Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$ is minimal with minimality index at most ℓ. Then, the extended pair $(\hat{X}, \hat{\Lambda})$ defined in* (6.1) *is minimal if and only if*

$$\mathbf{v}^p(y, z, \theta) \notin \operatorname{span} \mathbf{V}^p_{\ell+1}(X, \Lambda). \tag{6.5}$$

In particular, if the extended pair $(\hat{X}, \hat{\Lambda})$ is minimal, its minimality index cannot exceed $\ell + 1$.

Proof. Lemma 6.1.3 implies that $\mathbf{V}^p_{\ell+1}(\hat{X}, \hat{\Lambda}) = \big[\mathbf{V}^p_{\ell+1}(X, \Lambda),\ \mathbf{v}^p(y,\, z,\, \theta)\big]$, where $\mathbf{V}^p_{\ell+1}(X, \Lambda)$ has full column rank m thanks to the minimality of (X, Λ). Therefore, if (6.5) holds, $\mathbf{V}^p_{\ell+1}(\hat{X}, \hat{\Lambda})$ has full column rank $m + 1$, implying that the extended pair $(\hat{X}, \hat{\Lambda})$ is minimal with minimality index at most $\ell + 1$. If, on the other hand, (6.5) is violated, then $\operatorname{rank} \mathbf{V}^p_{\ell+1}(\hat{X}, \hat{\Lambda}) = m$. Because $\operatorname{rank} \mathbf{V}^p_{\ell}(\hat{X}, \hat{\Lambda}) \geq m$ by Lemma 6.1.3 and the minimality of (X, Λ), we infer from Proposition 3.1.6 that the extended pair $(\hat{X}, \hat{\Lambda})$ cannot be minimal. □

To enforce criterion (6.5) in a computational method, we impose the stronger condition

$$\mathbf{v}^p(y, z, \theta) \perp \operatorname{span} \mathbf{V}^p_{\ell+1}(X, \Lambda), \qquad \mathbf{v}^p(y, z, \theta) \neq 0.$$

The orthogonality requirement amounts to $\big[\mathbf{V}^p_{\ell+1}(X, \Lambda)\big]^{\mathsf{H}} \mathbf{v}^p(y, z, \theta) = 0$. Inserting the definitions of $\mathbf{V}^p_{\ell+1}(X, \Lambda)$ and $\mathbf{v}^p(y, z, \theta)$ in (3.10) and (6.4), respectively, this equation can be rewritten as

$$A(\theta)y + B(\theta)z = 0 \tag{6.6}$$

with the polynomials

$$\begin{aligned} A(\theta) &= p_0(\theta) \cdot p_0(\Lambda)^{\mathsf{H}} X^{\mathsf{H}} + \dots + p_\ell(\theta) \cdot p_\ell(\Lambda)^{\mathsf{H}} X^{\mathsf{H}},\\ B(\theta) &= p_0(\Lambda)^{\mathsf{H}} X^{\mathsf{H}} X q_0(\theta) + \dots + p_\ell(\Lambda)^{\mathsf{H}} X^{\mathsf{H}} X q_\ell(\theta). \end{aligned} \tag{6.7}$$

The non-degeneracy condition $\mathbf{v}^p(y, z, \theta) \neq 0$ is simplified by the upcoming lemma.

Lemma 6.1.5. *Let the pair (X, Λ) be minimal with minimality index at most ℓ and let $\mathbf{v}^p(y, z, \theta)$ be defined as in* (6.4)*. Then $\mathbf{v}^p(y, z, \theta) = 0$ if and only if $\left[\begin{smallmatrix} y\\ z\end{smallmatrix}\right] = 0$.*

Proof. We will first prove the lemma for the special case that the underlying polynomial basis is formed by the degree-graded, monic polynomials $\tilde{p}_i(\lambda) = (\lambda - \theta)^i$, $i = 0, \dots, \ell$. Applying Lemma 6.1.3 using this polynomial basis, one readily verifies that evaluating the corresponding polynomials $\tilde{q}_0, \dots, \tilde{q}_\ell$ at θ yields $\tilde{q}_0(\theta) = 0$ and $\tilde{q}_i(\theta) = (\Lambda - \theta I)^{i-1}$ for $i = 1, \dots, \ell$. Hence, we find

$$\mathbf{v}^{\tilde{p}}(y, z, \theta) = \begin{bmatrix} y\\ \mathbf{V}^{\tilde{p}}_\ell(X, \Lambda) \cdot z\end{bmatrix}$$

with $\mathbf{V}_\ell^{\tilde{p}}(X,\Lambda)$ as defined in (3.10). Since $\mathbf{V}_\ell^{\tilde{p}}(X,\Lambda)$ has full column rank thanks to the minimality of (X,Λ), this relation demonstrates that $\mathbf{v}^{\tilde{p}}(y,z,\theta) = 0$ and $\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right] = 0$ are equivalent.

To extend this result to an arbitrary polynomial basis $p_0,\dots,p_\ell$, we deduce from Proposition 3.2.1 that there exists a nonsingular matrix $P \otimes I$ such that $\mathbf{V}_{\ell+1}^{p}(\hat{X},\hat{\Lambda}) = (P \otimes I) \cdot \mathbf{V}_{\ell+1}^{\tilde{p}}(\hat{X},\hat{\Lambda})$. Equating only the last columns of the previous identity and applying Lemma 6.1.3 to both $\mathbf{V}_{\ell+1}^{p}(\hat{X},\hat{\Lambda})$ and $\mathbf{V}_{\ell+1}^{\tilde{p}}(\hat{X},\hat{\Lambda})$ leads to $\mathbf{v}^{p}(y,z,\theta) = (P \otimes I) \cdot \mathbf{v}^{\tilde{p}}(y,z,\theta)$. Consequently, $\mathbf{v}^{p}(y,z,\theta) = 0$ and $\mathbf{v}^{\tilde{p}}(y,z,\theta) = 0$ are equivalent, concluding the proof. □

Combining the invariance condition (6.2), the minimality condition (6.6), and the simplified non-degeneracy condition from Lemma 6.1.5, we obtain

$$\begin{bmatrix} T(\theta) & U(\theta) \\ A(\theta) & B(\theta) \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = 0, \qquad \begin{bmatrix} y \\ z \end{bmatrix} \neq 0. \tag{6.8}$$

Note that (6.8) again has the structure of a nonlinear eigenvalue problem. This nonlinear eigenvalue problem is of size $(n+m)\times(n+m)$, where $n \times n$ is the size of the original eigenvalue problem and m is the size of the existing minimal invariant pair $(X,\Lambda) \in \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times m}$. Since m can be expected to be quite small compared to n, the increase in size is only marginal.

By solving the nonlinear eigenvalue problem (6.8), we obtain $y \in \mathbb{C}^n$, $z \in \mathbb{C}^m$, and $\theta \in \mathcal{D}$ needed to expand the existing minimal invariant pair (X,Λ) via (6.1). Clearly, the matrix-valued function in (6.8) is holomorphic since this is true for each of its blocks: $T(\theta)$ is holomorphic by assumption, the holomorphy of $U(\theta)$ has been shown in Lemma 6.1.2, and $A(\theta)$, $B(\theta)$ are just polynomials. Thus, any technique for solving holomorphic nonlinear eigenvalue problems (see Section 1.2) can be applied to (6.8).

The subsequent theorem, which represents the main theoretical contribution of this chapter, states that by solving the eigenvalue problem (6.8) instead of (1.1), we indeed deflate the minimal invariant pair (X,Λ) from the computation.

Theorem 6.1.6. *Let (X,Λ) be a minimal invariant pair of the regular nonlinear eigenvalue problem* (1.1). *If $\left(\left[\begin{smallmatrix} Y \\ Z \end{smallmatrix}\right],\Theta\right)$ is a minimal invariant pair of the augmented nonlinear eigenvalue problem* (6.8)*, then $\left([X,Y], \left[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\right]\right)$ is a minimal invariant pair of the original nonlinear eigenvalue problem* (1.1). *Conversely, if $\left([X,Y], \left[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\right]\right)$ is a minimal invariant pair of the original nonlinear eigenvalue problem* (1.1)*, then there exists a unique matrix F such that $\left(\left[\begin{smallmatrix} Y-XF \\ Z-(\Lambda F - F\Theta) \end{smallmatrix}\right],\Theta\right)$ is a minimal invariant pair of the augmented nonlinear eigenvalue problem* (6.8). *In particular, if the eigenvalue problem* (1.1) *is regular, then the augmented eigenproblem* (6.8) *is regular as well.*

For the proof of the above theorem, it will be convenient to introduce some more notation and a number of intermediate results. We begin by defining block versions of $U(\theta)$, $A(\theta)$, and $B(\theta)$ in (6.3) and (6.7) as well as investigating their properties.

Definition 6.1.7. Let $p_0,\dots,p_\ell$ be the polynomial basis used for the definition of $A(\theta)$, $B(\theta)$ in (6.7) and let the polynomials $q_0,\dots,q_\ell$ be defined correspondingly as in Lemma 6.1.3. Furthermore, let $\mathcal{C}$ be a contour in $\mathcal{D}$, enclosing the eigenvalues of Θ in its interior. Then we define

$$\mathbf{U}(Z,\Theta) = \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}} U(\xi)Z(\xi I-\Theta)^{-1}\,\mathrm{d}\xi, \quad \mathbf{q}_i(Z,\Theta) = \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}} q_i(\xi)Z(\xi I-\Theta)^{-1}\,\mathrm{d}\xi,$$
$$\mathbf{A}(Y,\Theta) = \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}} A(\xi)Y(\xi I-\Theta)^{-1}\,\mathrm{d}\xi, \quad \mathbf{B}(Z,\Theta) = \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}} B(\xi)Z(\xi I-\Theta)^{-1}\,\mathrm{d}\xi.$$

Lemma 6.1.8. *For any $(Z,\Theta) \in \mathbb{C}^{m\times k}\times\mathbb{C}^{k\times k}$,*

$$\mathbf{U}(Z,\Theta) = \mathrm{D}_{[\Lambda,\Theta]}T(X,\cdot)Z.$$

Proof. Let $\mathcal{C}$ be a contour in $\mathcal{D}$ enclosing the eigenvalues of both Λ and Θ in its interior. Lemma 6.2.2 below implies that $U(\xi)(\xi I-\Lambda) = T(\xi)X$. Since all eigenvalues of Λ lie inside the contour $\mathcal{C}$, the factor $\xi I-\Lambda$ is nonsingular for all $\xi\in\mathcal{C}$. Therefore, solving this identity for $U(\xi)$ and inserting the result into the definition of $\mathbf{U}(Z,\Theta)$ implies

$$\mathbf{U}(Z,\Theta) = \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}} T(\xi)X(\xi I-\Lambda)^{-1}Z(\xi I-\Theta)^{-1}\,\mathrm{d}\xi.$$

The assertion now follows from Lemma 3.3.5. □

Lemma 6.1.9. *Let the polynomials $p_0,\dots,p_\ell$ as well as $q_0,\dots,q_\ell$ be defined as in Lemma 6.1.3. Then, for any $(Z,\Theta)\in\mathbb{C}^{m\times k}\times\mathbb{C}^{k\times k}$ and any $i=0,\dots,\ell$,*

$$\mathbf{q}_i(Z,\Theta) = (\mathrm{D}_{[\Lambda,\Theta]}p_i)Z.$$

Proof. Exploiting the linearity of the contour integral in the definition of $\mathbf{q}_i(Z,\Theta)$ as well as the Cauchy integral formula, one calculates that

$$\mathbf{q}_i(Z,\Theta) = \alpha_i\cdot\sum_{j=1}^{d_i}(\Lambda-\beta_{i,1}I)\cdots(\Lambda-\beta_{i,j-1}I)\cdot Z\cdot(\Theta-\beta_{i,j+1}I)\cdots(\Theta-\beta_{i,d_i}I).$$

Using this expression, an induction on the degree of the polynomial p_i shows

$$p_i\left(\begin{bmatrix}\Lambda & Z\\ 0 & \Theta\end{bmatrix}\right) = \begin{bmatrix}p_i(\Lambda) & \mathbf{q}_i(Z,\Theta)\\ 0 & p_i(\Theta)\end{bmatrix},$$

from which the assertion follows via Proposition 3.3.4. □

Employing the notation introduced above, we can now easily characterize the invariant pairs of the augmented nonlinear eigenvalue problem (6.8).

Lemma 6.1.10. *The pair $\left(\left[\begin{smallmatrix}Y\\Z\end{smallmatrix}\right],\Theta\right)$ is invariant with respect to the augmented nonlinear eigenvalue problem* (6.8) *if and only if the conditions*

$$\mathbf{T}(Y,\Theta)+\mathbf{U}(Z,\Theta)=0 \qquad \textit{and} \qquad \mathbf{A}(Y,\Theta)+\mathbf{B}(Z,\Theta)=0 \tag{6.9}$$

hold.

Proof. By Definition 3.1.1, $\left(\left[\begin{smallmatrix}Y\\Z\end{smallmatrix}\right],\Theta\right)$ constitutes an invariant pair of the augmented nonlinear eigenvalue problem (6.8) if and only if

$$0 = \frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}}\begin{bmatrix}T(\xi) & U(\xi)\\ A(\xi) & B(\xi)\end{bmatrix}\begin{bmatrix}Y\\ Z\end{bmatrix}(\xi I-\Theta)^{-1}\,\mathrm{d}\xi = \begin{bmatrix}\mathbf{T}(Y,\Theta)+\mathbf{U}(Z,\Theta)\\ \mathbf{A}(Y,\Theta)+\mathbf{B}(Z,\Theta)\end{bmatrix},$$

where the second equality is due to Definition 6.1.7. This equation obviously implies the claimed equivalence. □

Furthermore, we will need to characterize the invariance and minimality of augmented pairs of the form

$$\left([X,\,Y],\,\begin{bmatrix}\Lambda & Z\\ 0 & \Theta\end{bmatrix}\right) \in \mathbb{C}^{n\times(m+k)} \times \mathbb{C}^{(m+k)\times(m+k)}. \tag{6.10}$$

The subsequent results are devoted to this task.

Proposition 6.1.11. *Any pair of the form* (6.10) *satisfies*

$$\mathbf{T}\left([X,\,Y],\begin{bmatrix}\Lambda & Z\\ 0 & \Theta\end{bmatrix}\right) = \big[\mathbf{T}(X,\Lambda),\; \mathbf{T}(Y,\Theta) + \mathbf{U}(Z,\Theta)\big].$$

Proof. The claimed identity is an immediate consequence of Proposition 3.3.7 and Lemma 6.1.8. □

Proposition 6.1.12. *Let the polynomials $p_0,\dots,p_\ell$ constitute a basis for the vector space of polynomials of degree at most ℓ, and let the polynomials $q_0,\dots,q_\ell$ be defined correspondingly as in Lemma 6.1.3. Then, for any pair of the form* (6.10)*,*

$$\mathbf{V}^p_{\ell+1}\left([X,\,Y],\,\begin{bmatrix}\Lambda & Z\\ 0 & \Theta\end{bmatrix}\right) = \big[\mathbf{V}^p_{\ell+1}(X,\Lambda),\; \mathbf{v}^p(Y,Z,\Theta)\big], \tag{6.11}$$

where

$$\mathbf{v}^p(Y,Z,\Theta) = \begin{bmatrix} Yp_0(\Theta) + X\mathbf{q}_0(Z,\Theta)\\ \vdots \\ Yp_\ell(\Theta) + X\mathbf{q}_\ell(Z,\Theta)\end{bmatrix}. \tag{6.12}$$

Proof. The claimed identity follows by combining the statements of Lemma 3.3.6, Proposition 3.3.7, and Lemma 6.1.9. □

Lemma 6.1.13. *Let the polynomials $p_0,\dots,p_\ell$ constitute a basis for the vector space of polynomials of degree at most ℓ. Then, for any pair of the form* (6.10)*,*

$$\begin{aligned}&\mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}}\cdot\mathbf{V}^p_{\ell+1}\left([X,Y],\,\begin{bmatrix}\Lambda & Z\\ 0 & \Theta\end{bmatrix}\right)\\ &\quad= \Big[\mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}}\cdot\mathbf{V}^p_{\ell+1}(X,\Lambda),\; \mathbf{A}(Y,\Theta)+\mathbf{B}(Z,\Theta)\Big].\end{aligned}$$

Proof. Let the polynomials $q_0,\dots,q_\ell$ corresponding to $p_0,\dots,p_\ell$ be defined as in Lemma 6.1.3. From Definition 6.1.7, the linearity of the contour integral, and the Cauchy integral formula, we find

$$\begin{aligned}\mathbf{A}(Y,\Theta) &= p_0(\Lambda)^{\mathsf{H}}X^{\mathsf{H}}Yp_0(\Theta) + \cdots + p_\ell(\Lambda)^{\mathsf{H}}X^{\mathsf{H}}Yp_\ell(\Theta),\\ \mathbf{B}(Z,\Theta) &= p_0(\Lambda)^{\mathsf{H}}X^{\mathsf{H}}X\mathbf{q}_0(Z,\Theta) + \cdots + p_\ell(\Lambda)^{\mathsf{H}}X^{\mathsf{H}}X\mathbf{q}_\ell(Z,\Theta).\end{aligned}$$

Thus, by summing up,

$$\mathbf{A}(Y,\Theta)+\mathbf{B}(Z,\Theta) = \mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}}\cdot\mathbf{v}^p(Y,Z,\Theta)$$

with $\mathbf{V}^p_{\ell+1}(X,\Lambda)$ and $\mathbf{v}^p(Y,Z,\Theta)$ defined as in (3.10) and (6.12), respectively. Premultiplying Equation (6.11) obtained from Proposition 6.1.12 by $\mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}}$ and inserting the above relation completes the proof. □

Proof of Theorem 6.1.6. To establish the first statement of the theorem, suppose the pair $\big(\big[\begin{smallmatrix} Y \\ Z \end{smallmatrix}\big], \Theta\big)$ is minimal invariant with respect to the augmented nonlinear eigenvalue problem (6.8). Thus, by Lemma 6.1.10, the two conditions in (6.9) are satisfied. We first prove the invariance of the extended pair $\big([X, Y], \big[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\big]\big)$. According to Proposition 6.1.11, the block residual of the extended pair is given by

$$\mathbf{T}\left([X,\ Y],\ \begin{bmatrix} \Lambda & Z \\ 0 & \Theta \end{bmatrix}\right) = \big[\mathbf{T}(X, \Lambda),\ \mathbf{T}(Y, \Theta) + \mathbf{U}(Z, \Theta)\big].$$

The first condition in (6.9) shows that the second component on the right-hand side of the last equation is zero. The first component vanishes because (X, Λ) is invariant, implying the invariance of $\big([X, Y], \big[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\big]\big)$ as claimed.

To conclude that $\big([X, Y], \big[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\big]\big)$ is minimal, we have to demonstrate that for any $\lambda \in \mathbb{C}$, the equation

$$\begin{bmatrix} X & Y \\ \Lambda - \lambda I & Z \\ 0 & \Theta - \lambda I \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} = 0 \tag{6.13}$$

admits only the trivial solution $\big[\begin{smallmatrix} g \\ h \end{smallmatrix}\big] = 0$. Note that any solution of Equation (6.13) satisfies $[X, Y]\big[\begin{smallmatrix} g \\ h \end{smallmatrix}\big] = 0$ and $\big[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\big]\big[\begin{smallmatrix} g \\ h \end{smallmatrix}\big] = \lambda\big[\begin{smallmatrix} g \\ h \end{smallmatrix}\big]$. From this, we deduce the identities

$$[X,\ Y] \cdot p_i\left(\begin{bmatrix} \Lambda & Z \\ 0 & \Theta \end{bmatrix}\right) \begin{bmatrix} g \\ h \end{bmatrix} = p_i(\lambda) \cdot [X,\ Y] \begin{bmatrix} g \\ h \end{bmatrix} = 0, \qquad i = 0, \dots, \ell,$$

which when stacked combine to $\mathbf{V}^p_{\ell+1}\big([X, Y], \big[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\big]\big)\big[\begin{smallmatrix} g \\ h \end{smallmatrix}\big] = 0$. Premultiplying the last equation by $\mathbf{V}^p_{\ell+1}(X, \Lambda)^{\mathsf{H}}$ and applying Lemma 6.1.13, we obtain

$$\big[\mathbf{V}^p_{\ell+1}(X, \Lambda)^{\mathsf{H}} \cdot \mathbf{V}^p_{\ell+1}(X, \Lambda)\big] \cdot g + \big[\mathbf{A}(Y, \Theta) + \mathbf{B}(Z, \Theta)\big] \cdot h = 0.$$

Because of the second condition in (6.9), the summand involving h disappears, and the positive definiteness of the matrix $\mathbf{V}^p_{\ell+1}(X, \Lambda)^{\mathsf{H}} \cdot \mathbf{V}^p_{\ell+1}(X, \Lambda)$ entails $g = 0$. Hence, Equation (6.13) reduces to

$$\begin{bmatrix} Y \\ Z \\ \Theta - \lambda I \end{bmatrix} h = 0, \tag{6.14}$$

which has only the solution $h = 0$ as a result of the minimality of $\big(\big[\begin{smallmatrix} Y \\ Z \end{smallmatrix}\big], \Theta\big)$. This finishes the proof of the first statement.

For the converse statement, assume that $\big([X, Y], \big[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\big]\big)$ is a minimal invariant pair of the original nonlinear eigenvalue problem (1.1). This immediately implies the minimality of the pair $\big(\big[\begin{smallmatrix} Y \\ Z \end{smallmatrix}\big], \Theta\big)$ because for any solution h of Equation (6.14), $\big[\begin{smallmatrix} 0 \\ h \end{smallmatrix}\big]$ is a solution of Equation (6.13) and therefore zero.

It is clear from Lemma 6.1.10 that the pair $\big(\big[\begin{smallmatrix} Y \\ Z \end{smallmatrix}\big], \Theta\big)$ is invariant if and only if the two conditions in (6.9) hold. The first condition follows directly from the invariance of $\big([X, Y], \big[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\big]\big)$ by virtue of Proposition 6.1.11. The second condition, however, need not be satisfied in general.

Instead of $\big([X, Y], \big[\begin{smallmatrix} \Lambda & Z \\ 0 & \Theta \end{smallmatrix}\big]\big)$, we can also apply the above arguments starting from the transformed pair

$$\left([X,\, Y] \begin{bmatrix} I & F \\ 0 & I \end{bmatrix}^{-1},\; \begin{bmatrix} I & F \\ 0 & I \end{bmatrix} \begin{bmatrix} \Lambda & Z \\ 0 & \Theta \end{bmatrix} \begin{bmatrix} I & F \\ 0 & I \end{bmatrix}^{-1} \right) \\ = \left([X,\, Y - XF],\; \begin{bmatrix} \Lambda & Z - (\Lambda F - F\Theta) \\ 0 & \Theta \end{bmatrix} \right),$$

which is minimal and invariant by Lemma 3.2.3, regardless of the choice of the matrix F. This yields that the pair $\big(\big[\begin{smallmatrix} Y-XF \\ Z-(\Lambda F - F\Theta) \end{smallmatrix}\big], \Theta\big)$ is minimal and satisfies the first condition required for invariance. The second condition required for invariance of this pair reads

$$\mathbf{A}\big(Y - XF,\, \Theta\big) + \mathbf{B}\big(Z - (\Lambda F - F\Theta),\, \Theta\big) = 0, \tag{6.15}$$

and the proof of the converse statement is completed by determining F such that this condition is fulfilled. Applying twice Lemma 6.1.13 together with Proposition 3.2.3, we calculate

$$\begin{aligned} &\Big[\mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}} \cdot \mathbf{V}^p_{\ell+1}(X,\Lambda),\; \mathbf{A}\big(Y - XF,\, \Theta\big) + \mathbf{B}\big(Z - (\Lambda F - F\Theta),\, \Theta\big)\Big] \\ &= \mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}} \cdot \mathbf{V}^p_{\ell+1}\left([X,\, Y - XF],\; \begin{bmatrix} \Lambda & Z - (\Lambda F - F\Theta) \\ 0 & \Theta \end{bmatrix} \right) \\ &= \mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}} \cdot \mathbf{V}^p_{\ell+1}\left([X,\, Y],\; \begin{bmatrix} \Lambda & Z \\ 0 & \Theta \end{bmatrix} \right) \cdot \begin{bmatrix} I & F \\ 0 & I \end{bmatrix}^{-1} \\ &= \Big[\mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}} \cdot \mathbf{V}^p_{\ell+1}(X,\Lambda),\; \mathbf{A}(Y,\, \Theta) + \mathbf{B}(Z,\, \Theta)\Big] \cdot \begin{bmatrix} I & F \\ 0 & I \end{bmatrix}^{-1}. \end{aligned}$$

Equating only the second block components leads to

$$\begin{aligned} &\mathbf{A}\big(Y - XF,\, \Theta\big) + \mathbf{B}\big(Z - (\Lambda F - F\Theta),\, \Theta\big) \\ &\quad = \mathbf{A}(Y,\, \Theta) + \mathbf{B}(Z,\, \Theta) - \big[\mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}} \cdot \mathbf{V}^p_{\ell+1}(X,\Lambda)\big] F. \end{aligned}$$

The matrix $\mathbf{V}^p_{\ell+1}(X,\Lambda)^{\mathsf{H}} \cdot \mathbf{V}^p_{\ell+1}(X,\Lambda)$ is positive definite thanks to the minimality of (X,Λ). Hence, the above identity confirms the existence of a unique matrix F such that the second invariance condition in (6.15) is satisfied.

Finally, we prove the statement about the regularity by contradiction. To this end, assume that the augmented problem is singular; i.e., it has an eigenpair or, in other words, a minimal invariant pair $\big(\big[\begin{smallmatrix} y \\ z \end{smallmatrix}\big], \theta\big)$ for every $\theta \in \mathcal{D}$. As shown before, the extended pair $(\hat{X}, \hat{\Lambda})$ in (6.1) is then minimal and invariant with respect to the original problem. Obviously, θ is an eigenvalue of $\hat{\Lambda}$. Let u be a corresponding eigenvector, then $(\hat{X}u, \theta)$ is an eigenpair of the original problem. Since $\theta \in \mathcal{D}$ can be chosen arbitrarily, the original problem must be singular, in contradiction to the hypothesis. □

6.2 Algorithmic realization

In the following, we will derive an algorithm to efficiently solve the augmented nonlinear eigenvalue problems of the form (6.8) arising from the deflation strategy in the previous section. We begin by constructing a (simplified) Newton method and then turn this method into a Jacobi-Davidson-type algorithm by adding subspace acceleration as well as inexact solves of the correction equation. In this context, it will again be more convenient to consider the first and second block row of (6.8) individually.

6.2.1 A Newton approach. Suppose we already have an approximate solution $(y, z, \theta) \in \mathbb{C}^n \times \mathbb{C}^m \times \mathcal{D}$ of the augmented nonlinear eigenvalue problem (6.8) and want to compute a correction $(\triangle y, \triangle z, \triangle\theta)$ such that $(y+\triangle y, z+\triangle z, \theta+\triangle\theta)$ is an even better approximation. Ideally, the update leads to the exact solution, i.e.,

$$\begin{bmatrix} T(\theta+\triangle\theta) & U(\theta+\triangle\theta) \\ A(\theta+\triangle\theta) & B(\theta+\triangle\theta) \end{bmatrix} \begin{bmatrix} y+\triangle y \\ z+\triangle z \end{bmatrix} = 0. \tag{6.16}$$

To avoid the degenerate solution $(\triangle y,\, \triangle z) = (-y,\, -z)$, we additionally impose the orthogonality constraint

$$\begin{bmatrix} y \\ z \end{bmatrix}^{\mathsf{H}} \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} = 0. \tag{6.17}$$

Let σ be a shift close to the target eigenvalue. Employing Taylor expansion and neglecting higher order terms, (6.16) becomes

$$\begin{bmatrix} r \\ s \end{bmatrix} + \triangle\theta \begin{bmatrix} \dot{r} \\ \dot{s} \end{bmatrix} + \begin{bmatrix} T(\sigma) & U(\sigma) \\ A(\sigma) & B(\sigma) \end{bmatrix} \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} + (\theta + \triangle\theta - \sigma) \begin{bmatrix} \dot{T}(\sigma) & \dot{U}(\sigma) \\ \dot{A}(\sigma) & \dot{B}(\sigma) \end{bmatrix} \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} = 0, \tag{6.18}$$

where $\dot{T}$, $\dot{U}$, $\dot{A}$, $\dot{B}$ denote the derivatives with respect to θ of T, U, A, B, respectively, and

$$\begin{aligned} r &= T(\theta)y + U(\theta)z, & \dot{r} &= \dot{T}(\theta)y + \dot{U}(\theta)z, \\ s &= A(\theta)y + B(\theta)z, & \dot{s} &= \dot{A}(\theta)y + \dot{B}(\theta)z. \end{aligned} \tag{6.19}$$

If σ and θ are close asymptotically, the last summand on the left-hand side of (6.18) will be small. In fact, this term will be of second order if $\sigma - \theta = O(\triangle\theta)$. Neglecting the term and combining the remainder with the orthogonality condition (6.17) finally yields the linear system

$$\begin{bmatrix} T(\sigma) & U(\sigma) & \dot{r} \\ A(\sigma) & B(\sigma) & \dot{s} \\ y^{\mathsf{H}} & z^{\mathsf{H}} & 0 \end{bmatrix} \begin{bmatrix} \triangle y \\ \triangle z \\ \triangle\theta \end{bmatrix} = - \begin{bmatrix} r \\ s \\ 0 \end{bmatrix} \tag{6.20}$$

for computing the desired update.

By iteratively correcting an initial guess, we obtain Algorithm 6.1 to solve the augmented nonlinear eigenvalue problem (6.8). Local convergence of this algorithm towards simple eigenpairs can be proven using standard results [109, Theorem 4.1] on the convergence of simplified Newton methods. If the shift σ is updated in every step with the current eigenvalue approximation, Algorithm 6.1 is equivalent to nonlinear inverse iteration [112]. However, unlike nonlinear inverse iteration, keeping the shift constant to save computational work does not lead to erratic convergence. The convergence behavior for multiple eigenvalues is harder to analyze; see [68, 129] for convergence analyses of related methods in the presence of multiple eigenvalues. It is, however, easily seen from Theorem 6.1.6 that any multiple eigenvalue of the augmented eigenvalue problem (6.8) is a multiple

Algorithm 6.1: A simplified Newton method for solving the augmented nonlinear eigenvalue problem (6.8).

```
Input: minimal invariant pair $(X, \Lambda)$, initial approximation $(y_0, z_0, \theta_0)$
Output: solution $(y, z, \theta)$ of (6.8)

determine minimality index $\ell$ of $(X, \Lambda)$
for $k = 0, 1, 2, \dots$ until convergence do
    pick a shift $\sigma = \sigma_k$
    solve the correction equation (6.20) for $\triangle y$, $\triangle z$, $\triangle\theta$
    $y_{k+1} = y_k + \triangle y$
    $z_{k+1} = z_k + \triangle z$
    $\theta_{k+1} = \theta_k + \triangle\theta$
end
```

eigenvalue of the original eigenproblem (1.1) as well. Hence, the difficulties are not inherent to the deflation approach in Section 6.1.

Once the algorithm has found a solution of the augmented nonlinear eigenvalue problem (6.8), the current minimal invariant pair (X, Λ) is expanded via (6.1). According to Lemmas 6.1.1 and 6.1.4, the resulting pair $(\hat{X}, \hat{\Lambda})$ is again invariant and minimal.

6.2.2 A Jacobi-Davidson-type algorithm. Instead of directly updating the current iterate as in the previous subsection, the correction computed from the linear system (6.20) can also be used to expand the search space in a Petrov-Galerkin projection framework for solving the augmented nonlinear eigenvalue problem (6.8). In this case, we are only interested in the $\triangle y$- and $\triangle z$-components of the solution. Hence, we eliminate $\triangle\theta$ from the system as follows. First, we recast the system as

$$\begin{bmatrix} T(\sigma) & U(\sigma) \\ A(\sigma) & B(\sigma) \end{bmatrix} \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} = - \begin{bmatrix} r + \dot{r}\triangle\theta \\ s + \dot{s}\triangle\theta \end{bmatrix}, \quad \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} \perp \begin{bmatrix} y \\ z \end{bmatrix}.$$

Next, we premultiply by the oblique projector $I - \left[\begin{smallmatrix} \dot{r} \\ \dot{s} \end{smallmatrix}\right] w^{\mathsf{H}}$ with the vector $w \in \mathbb{C}^{n+m}$ chosen orthogonal to $\left[\begin{smallmatrix} r \\ s \end{smallmatrix}\right]$ and normalized such that $w^{\mathsf{H}} \left[\begin{smallmatrix} \dot{r} \\ \dot{s} \end{smallmatrix}\right] = 1$. This gives

$$\left(I - \begin{bmatrix} \dot{r} \\ \dot{s} \end{bmatrix} w^{\mathsf{H}} \right) \begin{bmatrix} T(\sigma) & U(\sigma) \\ A(\sigma) & B(\sigma) \end{bmatrix} \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} = - \begin{bmatrix} r \\ s \end{bmatrix}, \quad \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} \perp \begin{bmatrix} y \\ z \end{bmatrix}.$$

Because of the orthogonality condition, the last equation can also be written as

$$\left(I - \begin{bmatrix} \dot{r} \\ \dot{s} \end{bmatrix} w^{\mathsf{H}} \right) \begin{bmatrix} T(\sigma) & U(\sigma) \\ A(\sigma) & B(\sigma) \end{bmatrix} \left(I - \begin{bmatrix} y \\ z \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}^{\mathsf{H}} \right) \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} = - \begin{bmatrix} r \\ s \end{bmatrix}, \quad \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} \perp \begin{bmatrix} y \\ z \end{bmatrix}, \tag{6.21}$$

assuming, without loss of generality, $\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right]^{\mathsf{H}} \left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right] = 1$. Equation (6.21) has the form of a Jacobi-Davidson correction equation, similar to the work in [137] but with more freedom in the choice of w.

An algorithm for the solution of the augmented nonlinear eigenvalue problem (6.8) based on (6.21) would proceed as follows. Suppose $\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right] \in \mathbb{C}^{(n+m)\times k}$ is a matrix having orthonormal columns which span the current search space. The

Algorithm 6.2: Nonlinear Jacobi-Davidson algorithm with deflation.

```
Input: minimal invariant pair $(X, \Lambda)$, initial approximation $(y_0, z_0, \theta_0)$, basis
       for initial search space $\left[\begin{smallmatrix} Y_0 \\ Z_0 \end{smallmatrix}\right]$, basis for initial test space $\left[\begin{smallmatrix} W_{1,0} \\ W_{2,0} \end{smallmatrix}\right]$
Output: extended minimal invariant pair $(\hat{X}, \hat{\Lambda})$

determine minimality index $\ell$ of $(X, \Lambda)$
for $k = 0, 1, 2, \dots$ do
    compute residual $\left[\begin{smallmatrix} r \\ s \end{smallmatrix}\right]$ and derivative of residual $\left[\begin{smallmatrix} \dot{r} \\ \dot{s} \end{smallmatrix}\right]$ as defined in (6.19)
    if residual norm below convergence threshold then
        build $(\hat{X}, \hat{\Lambda})$ via (6.1), stop.
    end
    (approximately) solve correction equation (6.21) for $\left[\begin{smallmatrix} \triangle y \\ \triangle z \end{smallmatrix}\right]$
    expand search space $\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right]$ to $\left[\begin{smallmatrix} Y_{k+1} \\ Z_{k+1} \end{smallmatrix}\right]$ as in (6.23)
    expand test space $\left[\begin{smallmatrix} W_{1,k} \\ W_{2,k} \end{smallmatrix}\right]$ to $\left[\begin{smallmatrix} W_{1,k+1} \\ W_{2,k+1} \end{smallmatrix}\right]$ as in (6.24)
    solve projected eigenvalue problem by contour integral method
    if no eigenvalues found then
        perform Newton update $\left[\begin{smallmatrix} y_{k+1} \\ z_{k+1} \end{smallmatrix}\right] = \left[\begin{smallmatrix} y_k \\ z_k \end{smallmatrix}\right] + \left[\begin{smallmatrix} \triangle y \\ \triangle z \end{smallmatrix}\right]$
        set $\theta_{k+1} = \theta_k$
    else
        choose eigenpair $(u, \theta)$ of projected problem with $\theta$ closest to $\theta_k$
        set $\theta_{k+1} = \theta$, $\left[\begin{smallmatrix} y_{k+1} \\ z_{k+1} \end{smallmatrix}\right] = \left[\begin{smallmatrix} Y_{k+1} \\ Z_{k+1} \end{smallmatrix}\right] u$
    end
end
```

current eigenpair approximation $\big(\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right], \theta\big)$ is then given by $\big(\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right] u, \theta\big)$, where (u, θ) with $u^{\mathsf{H}} u = 1$ is an eigenpair of the projected nonlinear eigenproblem

$$\begin{bmatrix} W_{1,k} \\ W_{2,k} \end{bmatrix}^{\mathsf{H}} \begin{bmatrix} T(\theta) & U(\theta) \\ A(\theta) & B(\theta) \end{bmatrix} \begin{bmatrix} Y_k \\ Z_k \end{bmatrix} u = 0 \tag{6.22}$$

for some matrix $\left[\begin{smallmatrix} W_{1,k} \\ W_{2,k} \end{smallmatrix}\right] \in \mathbb{C}^{(n+m) \times k}$ with orthonormal columns. The eigenpair of the projected problem should be selected such that θ is as close as possible to the target eigenvalue of (6.8). Now the correction equation (6.21) is solved for $\triangle y$, $\triangle z$ and $\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right]$ is expanded to $\left[\begin{smallmatrix} Y_{k+1} \\ Z_{k+1} \end{smallmatrix}\right]$ having orthonormal columns such that

$$\operatorname{span} \begin{bmatrix} Y_{k+1} \\ Z_{k+1} \end{bmatrix} = \operatorname{span} \left\{ \begin{bmatrix} Y_k \\ Z_k \end{bmatrix}, \begin{bmatrix} \triangle y \\ \triangle z \end{bmatrix} \right\}. \tag{6.23}$$

The entire procedure is repeated with $\left[\begin{smallmatrix} Y_{k+1} \\ Z_{k+1} \end{smallmatrix}\right]$ in place of $\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right]$ until the desired accuracy of the approximate eigenpair is reached. Afterwards, the computed (y, z, θ) is used to expand the current invariant pair via (6.1). As in the previous section, the shift σ may be updated periodically to speed up convergence or kept constant to save computational work. The above framework is summarized in Algorithm 6.2. In the following, we comment on the details of a practical implementation.

6.2.3 Choice of search and test spaces. If available, the search space $\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right]$ should be initialized with a normalized approximation $\left[\begin{smallmatrix} \tilde{y} \\ \tilde{z} \end{smallmatrix}\right]$ to the desired eigenvector of the augmented nonlinear eigenvalue problem (6.8). In case only a normalized, approximate eigenvector $\tilde{y}$ for the original nonlinear eigenproblem (1.1) is known, an adequate initialization is given by $\left[\begin{smallmatrix}\tilde{y}\\0\end{smallmatrix}\right]$. In the absence of any suitable approximations, the search space is initialized with a normalized random vector. The test space $\left[\begin{smallmatrix}W_{1,k}\\W_{2,k}\end{smallmatrix}\right]$ is initialized with a normalized version of $\left[\begin{smallmatrix}\dot{r}\\\dot{s}\end{smallmatrix}\right]$ computed from the initial search space in the first step of the algorithm and then expanded in every iteration to include the current residual. That is, $\left[\begin{smallmatrix}W_{1,k+1}\\W_{2,k+1}\end{smallmatrix}\right]$ is chosen such that it has orthonormal columns and

$$\operatorname{span}\begin{bmatrix}W_{1,k+1}\\W_{2,k+1}\end{bmatrix}=\operatorname{span}\left\{\begin{bmatrix}W_{1,k}\\W_{2,k}\end{bmatrix},\begin{bmatrix}r\\s\end{bmatrix}\right\}. \tag{6.24}$$

This strategy can be viewed as a generalization of the harmonic Rayleigh-Ritz extraction procedure [107], which is frequently employed in connection with the Jacobi-Davidson algorithm for linear eigenvalue problems [119, 120].

6.2.4 Solution of the projected eigenproblems. For solving the projected eigenproblems we employ the contour integral method [19, 7], followed by a few steps of Newton-based iterative refinement. As contour we choose a circle with a prescribed radius around the current eigenvalue approximation. Especially during the first steps of the algorithm, it may happen that there are no eigenvalues of the projected eigenproblem inside the contour. In this event, we use the computed solution of the correction equation (6.21) to update the eigenvector approximation to $\left[\begin{smallmatrix}y+\triangle y\\z+\triangle z\end{smallmatrix}\right]$ as in the Newton method described in Section 6.2.1 while leaving the eigenvalue approximation unchanged.

In principle, any other solution method for nonlinear eigenvalue problems could be used as well to handle the projected problems. Most notably, when dealing with polynomial eigenvalue problems, the augmented problem (6.8) will be polynomial as well, facilitating the use of linearization techniques [48, 93]. Ideally, the method of choice should be able to benefit significantly from the small size of the projected problems.

6.2.5 Solution of the correction equation. As is typical for a Jacobi-Davidson iteration, the correction equation (6.21) need not be solved very accurately. A few steps of a preconditioned Krylov subspace method, such as GMRES, are usually sufficient. In our experiments, we have tested stopping the Krylov solver after having decreased the residual by a prescribed factor as well as stopping the solver after a fixed number of iterations. Both strategies seem to work equally well.

To enable the iterative solution of the correction equation (6.21), we are in need of an effective preconditioner. In the following, we will describe how a suitable preconditioner can be constructed based on an existing preconditioner for $T(\theta)$. Since devising a good preconditioner for $T(\theta)$ is a highly application-dependent task, we will assume here that it is given to us.

Recalling that the blocks within the matrix

$$\mathcal{T}=\begin{bmatrix}T & U\\A & B\end{bmatrix} \tag{6.25}$$

associated with the augmented nonlinear eigenvalue problem (6.8) have different origins, it seems wise to take the block structure into account when thinking about preconditioning. Note that we have dropped the explicit dependence on θ here for the sake of better readability. A rich theory on preconditioning of 2×2 block systems is available in the literature; see [14] as well as the references therein. Unfortunately, most of the existing methods require the blocks to possess additional properties, such as symmetry or definiteness, which, in general, are not present in $\mathcal{T}$. For this reason, we will pursue a different approach.

Consider the block triangular factorization

$$\begin{bmatrix} T & U \\ A & B \end{bmatrix} = \begin{bmatrix} I & 0 \\ AT^{-1} & I \end{bmatrix} \begin{bmatrix} T & U \\ 0 & B - AT^{-1}U \end{bmatrix}$$

of $\mathcal{T}$. The inverse of the right block triangular factor reads

$$\begin{bmatrix} T^{-1} & -T^{-1}U(B - AT^{-1}U)^{-1} \\ 0 & (B - AT^{-1}U)^{-1} \end{bmatrix}.$$

Approximating the upper, left block of this inverse by a given preconditioner P^{-1} for T yields the matrix

$$\mathcal{P}^{-1} = \begin{bmatrix} P^{-1} & -T^{-1}U(B - AT^{-1}U)^{-1} \\ 0 & (B - AT^{-1}U)^{-1} \end{bmatrix}, \tag{6.26}$$

which we will employ as a preconditioner for $\mathcal{T}$. The following proposition is an extension of [63, Proposition 1].

Proposition 6.2.1. *Assume that the matrix $\mathcal{T}$ in* (6.25) *is preconditioned by $\mathcal{P}$ as in* (6.26). *Then the spectrum of the preconditioned matrix $\mathcal{T}\mathcal{P}^{-1}$ consists of the spectrum of TP^{-1} and the eigenvalue* 1. *Furthermore, the degree of the minimal polynomial of $\mathcal{T}\mathcal{P}^{-1}$ exceeds the degree of the minimal polynomial of TP^{-1} by at most one.*

Proof. The statement about the spectrum is obvious from the structure of the preconditioned matrix,

$$\mathcal{T}\mathcal{P}^{-1} = \begin{bmatrix} TP^{-1} & 0 \\ AP^{-1} & I \end{bmatrix}.$$

For the second statement, let p be the minimal polynomial of TP^{-1}. One easily calculates that $(\mathcal{T}\mathcal{P}^{-1} - I) \cdot p(\mathcal{T}\mathcal{P}^{-1}) = 0$. Consequently, the minimal polynomial of $\mathcal{T}\mathcal{P}^{-1}$ must be a divisor of $(\lambda - 1) \cdot p(\lambda)$, implying the claimed bound on its degree. □

Proposition 6.2.1 demonstrates that the preconditioner $\mathcal{P}^{-1}$ for $\mathcal{T}$ has about the same quality as the preconditioner P^{-1} for T from which it has been constructed. However, it may seem that one needs to know T^{-1} in order to apply $\mathcal{P}^{-1}$. To eliminate this flaw, we utilize the subsequent lemma.

Lemma 6.2.2. *Let (X, Λ) be an invariant pair of the nonlinear eigenvalue problem* (1.1) *and let U be defined as in* (6.3). *Then for any $\theta \in \mathcal{D}$,*

$$U(\theta)(\theta I - \Lambda) = T(\theta)X.$$

Proof. A direct calculation using $\theta I - \Lambda = (\xi I - \Lambda) - (\xi - \theta)I$ shows

$$\begin{aligned} U(\theta)(\theta I - \Lambda) &= \frac{1}{2\pi \mathrm{i}} \int_\Gamma T(\xi)X(\xi - \theta)^{-1} \,\mathrm{d}\xi - \frac{1}{2\pi \mathrm{i}} \int_\Gamma T(\xi)X(\xi I - \Lambda)^{-1} \,\mathrm{d}\xi \\ &= T(\theta)X - \mathbf{T}(X, \Lambda). \end{aligned}$$

The proof is finished by concluding from Definition 3.1.1 that $\mathbf{T}(X, \Lambda) = 0$. □

Suppose that θ is not an eigenvalue of T. Note that this also implies that θ is not an eigenvalue of Λ if the pair (X, Λ) is minimal invariant. Then Lemma 6.2.2 yields $T(\theta)^{-1}U(\theta) = X(\theta I - \Lambda)^{-1}$. However, if θ is close to an eigenvalue of Λ, the matrix $\theta I - \Lambda$ may be arbitrarily ill-conditioned. We therefore utilize $T(\theta)^{-1}U(\theta) = X(\theta I - \Lambda)^{\dagger}$ instead. Because the size of the matrix $\theta I - \Lambda$ can be expected to be rather small, the Moore-Penrose pseudo-inverse $(\theta I - \Lambda)^{\dagger}$ is efficiently computable by means of the singular value decomposition. Thus, the preconditioner applied in practice reads

$$\tilde{\mathcal{P}}^{-1} = \begin{bmatrix} P^{-1} & -X(\theta I - \Lambda)^{\dagger}\big(B - AX(\theta I - \Lambda)^{\dagger}\big)^{-1} \\ 0 & \big(B - AX(\theta I - \Lambda)^{\dagger}\big)^{-1} \end{bmatrix}.$$

Its application to a vector requires one linear system solve with P and the Schur complement $B - AX(\theta I - \Lambda)^{\dagger}$ each, as well as one matrix-vector multiplication by $X(\theta I - \Lambda)^{\dagger}$. Since the Schur complement is as small as the matrix $\theta I - \Lambda$, it can be inverted by a direct solver at negligible cost. Consequently, the computational work for applying $\tilde{\mathcal{P}}^{-1}$ is essentially that of applying P^{-1}.

In the Jacobi-Davidson correction equation (6.21), the matrix $\mathcal{T}$ is surrounded by projectors, which restrict its action to a map from the orthogonal complement of $\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right]$ to the orthogonal complement of w. The same restriction should also be applied to the preconditioner [120, 137]. It is straightforward to compute that the appropriately projected preconditioner is given by

$$\left(I - \frac{\tilde{\mathcal{P}}^{-1}\left[\begin{smallmatrix} \dot{r} \\ \dot{s} \end{smallmatrix}\right]\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right]^{\mathsf{H}}}{\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right]^{\mathsf{H}}\tilde{\mathcal{P}}^{-1}\left[\begin{smallmatrix} \dot{r} \\ \dot{s} \end{smallmatrix}\right]} \right)\tilde{\mathcal{P}}^{-1}.$$

Every application of the projector costs just one inner product. Additionally, we have to apply the preconditioner $\tilde{\mathcal{P}}^{-1}$ to $\left[\begin{smallmatrix} \dot{r} \\ \dot{s} \end{smallmatrix}\right]$ and take the inner product with $\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right]$ once in a preprocessing step.

6.2.6 Evaluating the residuals and projected problems. In every step of Algorithm 6.2, we have to evaluate the residual $\left[\begin{smallmatrix} r \\ s \end{smallmatrix}\right]$ along with its derivative $\left[\begin{smallmatrix} \dot{r} \\ \dot{s} \end{smallmatrix}\right]$ as defined in (6.19). A similar calculation as in the proof of Lemma 6.1.1 shows that r and $\dot{r}$ can be computed at once from the block residual

$$\mathbf{T}\left(\big[X,\ y,\ 0\big], \begin{bmatrix} \Lambda & z & \\ & \theta & 1 \\ & & \theta \end{bmatrix} \right) = \big[\mathbf{T}(X, \Lambda),\ r,\ \dot{r}\big]$$

of the original nonlinear eigenproblem (1.1). From the definitions of $A(\theta)$ and $B(\theta)$ in (6.7), it is evident that evaluating s and $\dot{s}$ involves mainly operations on small matrices and vectors, the only exceptions being the computation of $X^{\mathsf{H}}y$ and $X^{\mathsf{H}}X$. The latter, however, can be precomputed incrementally as X is built up. Using the calculus of matrix functions [60], the necessary polynomial evaluations can be conveniently combined into computing the matrix polynomials

$$p_i\left(\begin{bmatrix} \Lambda & z & \\ & \theta & 1 \\ & & \theta \end{bmatrix} \right) = \begin{bmatrix} p_i(\Lambda) & q_i(\theta)z & \dot{q}_i(\theta)z \\ & p_i(\theta) & \dot{p}_i(\theta) \\ & & p_i(\theta) \end{bmatrix}, \qquad i = 0, \dots, \ell.$$

The evaluation of the matrices associated with the projected nonlinear eigenvalue problems (6.22) proceeds in a similar fashion as above. If the block residual admits a representation of the form (3.2), the computational effort can be reduced by working with the projected coefficient matrices

$$W_{1,k}^{\mathsf{H}} \cdot T_j \cdot \big[X,\, Y_k\big], \qquad j = 1, \dots, d.$$

Again, it is beneficial to incrementally build up these projected coefficient matrices as we expand X, or the search and test spaces.

6.2.7 Restarting, Locking and Purging. The Jacobi-Davidson algorithm 6.2 is designed to aim at one eigenpair at a time. However, the search space may also contain approximations to eigenvectors whose corresponding eigenvalue lies close to the current target and it is desirable to use this information. On the other hand, we must not let the search space grow too large due to memory constraints. To keep the size of the search space at a moderate level, it needs to be purged periodically, i.e., replaced by a subspace of smaller dimension, which (hopefully) still contains the most valuable information. Since purging impedes convergence, we purge the search space only immediately after a target eigenpair has converged.

Suppose that convergence occurs at the k-th step of Algorithm 6.2. If the current search space $\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right]$ contains sufficient information about further eigenpairs, approximations to these eigenpairs will show up as solutions of the projected nonlinear eigenvalue problem (6.22). These approximations will be discovered by the contour integral method used to solve (6.22), provided that the associated eigenvalues lie close enough to the target. We will thus obtain a minimal invariant pair (C, Θ) of the projected eigenproblem (6.22), representing all eigenvalues inside the contour. By lifting this pair to the full space, $\big(\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right] C, \Theta\big)$ is an approximate minimal invariant pair of the augmented nonlinear eigenvalue problem (6.8). Hence, in compressing the search space to the range of $\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right] C$, we retain only the information necessary to reproduce the eigenpairs encoded by (C, Θ). If (C, Θ) consists only of the target eigenpair, we take this as an indication that no more relevant information is present in $\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right]$. In this event, the search space is purged entirely and replaced by a normalized, random vector unless further targets are left from previous restarts. Likewise, the contour integral method supplies information about the left eigenvectors, which can be used analogously to compress the test space $\left[\begin{smallmatrix} W_{1,k} \\ W_{2,k} \end{smallmatrix}\right]$.

Unfortunately, we cannot simply continue Algorithm 6.2 with the compressed search and test spaces because locking the currently converged eigenpair via (6.1) will increase the dimension of the augmented nonlinear eigenproblem (6.8) by one. Consequently, the search and test spaces need to be adjusted accordingly. We begin by describing this adjustment for the search space. Without loss of generality, we can assume Θ to be in Schur form, ordered in such a way that the first diagonal entry of Θ is the target eigenvalue and the second diagonal entry is the eigenvalue approximation we would like to select as the next target. Otherwise, we compute a unitary matrix Q such that $Q^{\mathsf{H}} \Theta Q$ has the desired form and work with the pair $(CQ, Q^{\mathsf{H}} \Theta Q)$ instead; compare Lemma 3.2.3. Partitioning

$$\begin{bmatrix} Y_k \\ Z_k \end{bmatrix} C = \begin{bmatrix} y_1 & Y_2 \\ z_1 & Z_2 \end{bmatrix}, \qquad \Theta = \begin{bmatrix} \theta & \Theta_{12} \\ 0 & \Theta_{22} \end{bmatrix}$$

such that $y_1 \in \mathbb{C}^n$ and $z_1 \in \mathbb{C}^m$ are vectors and θ is a scalar, $\big(\left[\begin{smallmatrix} y_1 \\ z_1 \end{smallmatrix}\right], \theta\big)$ is the eigenpair of the augmented eigenproblem (6.8) which is going to be locked. If the pair $\big(\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right] C, \Theta\big)$ were an exact minimal invariant pair of the augmented problem (6.8), then, by Theorem 6.1.6,

$$\left(\begin{bmatrix} X, & y_1, & Y_2 \end{bmatrix}, \begin{bmatrix} \Lambda & z_1 & Z_2 \\ 0 & \theta & \Theta_{12} \\ 0 & 0 & \Theta_{22} \end{bmatrix} \right)$$

would represent a minimal invariant pair of the original nonlinear eigenvalue problem (1.1). Locking the eigenpair $\left(\left[\begin{smallmatrix} y_1 \\ z_1 \end{smallmatrix} \right], \theta \right)$ and applying Theorem 6.1.6 again, we find that for a suitably chosen matrix F,

$$\left(\begin{bmatrix} Y_2 - \begin{bmatrix} X & y_1 \end{bmatrix} F \\ \begin{bmatrix} Z_2 \\ \Theta_{12} \end{bmatrix} - \left(\begin{bmatrix} \Lambda & z_1 \\ 0 & \theta \end{bmatrix} F - F\Theta_{22} \right) \end{bmatrix}, \Theta_{22} \right)$$

is a minimal invariant pair of the new augmented eigenproblem (6.8). Applying the same transformation also in the situation where $\left(\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix} \right], \Theta \right)$ constitutes only an approximate minimal invariant pair seems to be a reasonable heuristic. Recall from the proof of Theorem 6.1.6 that the choice of F only influences the minimality constraint (6.6) but neither the invariance constraint (6.2) nor the minimality of the ensuing pair itself. Moreover, we have not observed a significant gain from choosing the correct F in our experiments. Therefore, we save some computational work by setting $F = 0$. Performing an economy-size QR decomposition,

$$\begin{bmatrix} Y_2 \\ Z_2 \\ \Theta_{12} \end{bmatrix} = QR,$$

the resulting pair is transformed to $(Q, R\Theta_{22}R^{-1})$ according to Lemma 3.2.3. We then take Q as basis for the new search space and the first diagonal element of $R\Theta_{22}R^{-1}$ as the new shift. If $(Q, R\Theta_{22}R^{-1})$ contains more than one eigenpair, the unused ones are stored for future restarts.

Transforming the test space in a similar way does not make sense because the minimality conditions (6.6) before and after the locking are very different. Instead we propose partitioning the compressed test space as $\left[\begin{smallmatrix} W_1 \\ W_2 \end{smallmatrix} \right]$ conformally with the block structure of the augmented nonlinear eigenvalue problem (6.8) and then taking the new test space as the range of $\left[\begin{smallmatrix} W_1 \\ 0 \end{smallmatrix} \right]$.

6.2.8 Selective deflation. The computational effort per iteration of Algorithm 6.2 grows with the number of columns, m, of which the deflated minimal invariant pair $(X, \Lambda) \in \mathbb{C}^{n \times m} \times \mathbb{C}^{m \times m}$ consists. This effect, however, becomes severe only for large values of m. More dramatically, a larger invariant pair might have a higher minimality index, which is already indicated by the fact that the minimality index is bounded by m; see Corollary 3.1.7.

A higher minimality index impacts the evaluation of $A(\theta)$ and $B(\theta)$ defined in (6.7) in two ways. Firstly, it increases the necessary computational work by increasing the number of summands in (6.7). Besides and more importantly, it complicates the selection of a polynomial basis $p_0, \dots, p_\ell$ which avoids numerical instabilities in forming $A(\theta)$, $B(\theta)$ (see the discussion in Section 3.2) unless the eigenvalues of Λ are very well clustered.

It is therefore desirable to keep the size of the deflated minimal invariant pair as small as possible. The key to doing so lies in the observation that because Algorithm 6.2 operates only locally, reconvergence can only occur for eigenvalues which are close to the current target θ_0. In particular, if we solve the projected eigenproblems (6.22) by the contour integral method, where the contour is a circle of radius ρ around θ_0, only eigenvalues λ satisfying $|\lambda - \theta_0| \leq \rho$ are threatened by reconvergence. This motivates the following approach. We reorder (X, Λ) by a unitary transformation Q into an equivalent (in the sense of Lemma 3.2.3) pair

$$(XQ, Q^{\mathsf{H}}\Lambda Q) = \left([X_1, X_2],\, \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ & \Lambda_{22} \end{bmatrix} \right) \tag{6.27}$$

such that the eigenvalues of Λ_{11} and Λ_{22} lie inside and outside of the ball with radius $\gamma \cdot \rho$ around the target θ_0, respectively. Here, $\gamma > 1$ is a safety factor intended to compensate for possible updates of the target θ_0 within the course of the algorithm. The search and test spaces $\left[\begin{smallmatrix} Y_k \\ Z_k \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} W_{1,k} \\ W_{2,k} \end{smallmatrix}\right]$ have to be transformed accordingly to $\left[\begin{smallmatrix} Y_k \\ Q^{\mathsf{H}} Z_k \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} W_{1,k} \\ Q^{\mathsf{H}} W_{2,k} \end{smallmatrix}\right]$, respectively. We then construct the augmented nonlinear eigenvalue problem (6.8) only from (X_1, Λ_{11}). Since (X_1, Λ_{11}) tends to have fewer columns than the full pair (X, Λ), we have achieved our goal. Moreover, the eigenvalues of Λ_{11} are contained inside a ball. Thus, they tend to be better clustered than those of Λ, simplifying the choice of an appropriate polynomial basis.

While running Algorithm 6.2 to solve the augmented eigenproblem (6.8), we carefully monitor whether the eigenvalue approximation moves close to an eigenvalue of Λ_{22}, and if so, adjust the partitioning (6.27). This ensures that the algorithm computes an eigenpair $\left(\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right], \theta\right)$ such that θ is not within the spectrum of Λ_{22}. In lieu of the classical update given in (6.1), we then perform the expansion

$$(\hat{X}, \hat{\Lambda}) = \left([X_1,\ X_2,\ y],\, \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & z \\ & \Lambda_{22} & 0 \\ & & \theta \end{bmatrix} \right)$$

which is justified by the subsequent lemma.

Lemma 6.2.3. *Let both* $\left([X_1,\ X_2],\ \left[\begin{smallmatrix} \Lambda_{11} & \Lambda_{12} \\ & \Lambda_{22} \end{smallmatrix}\right]\right)$ *and* $\left([X_1,\ Y],\ \left[\begin{smallmatrix} \Lambda_{11} & Z \\ & \Theta \end{smallmatrix}\right]\right)$ *be minimal invariant pairs of the nonlinear eigenproblem* (1.1) *and assume that* Λ_{22} *and* Θ *have no eigenvalues in common. Then the extended pair*

$$\left([X_1,\ X_2,\ Y],\, \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & Z \\ & \Lambda_{22} & 0 \\ & & \Theta \end{bmatrix} \right)$$

is minimal and invariant.

Proof. The invariance follows by expanding and comparing the block residuals of the three pairs as in the proof of Lemma 6.1.1. To prove the minimality, we have to show that the equation

$$\begin{bmatrix} X_1 & X_2 & Y \\ \Lambda_{11} - \lambda I & \Lambda_{12} & Z \\ & \Lambda_{22} - \lambda I & 0 \\ & & \Theta - \lambda I \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ h \end{bmatrix} = 0$$

admits only the zero solution. Since Λ_{22} and Θ do not have any eigenvalues in common, at least one of the matrices $\Lambda_{22} - \lambda I$ and $\Theta - \lambda I$ is invertible. Without loss of generality, assume that the latter is invertible; otherwise, similar arguments apply. Then the last block row of the equation implies $h = 0$, leaving us with

$$\begin{bmatrix} X_1 & X_2 \\ \Lambda_{11} - \lambda I & \Lambda_{12} \\ & \Lambda_{22} - \lambda I \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = 0,$$

which only admits the zero solution thanks to the minimality of the first pair. □

Since Lemma 6.2.3 applies to blocks $\left(\left[\begin{smallmatrix} Y \\ Z \end{smallmatrix}\right], \Theta\right)$ and not just eigenpairs $\left(\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right], \theta\right)$, the restarting mechanism from Section 6.2.7 remains valid in the case of selective deflation.

6.3 Numerical experiments

To demonstrate the effectiveness of the deflation approach proposed in this chapter, we apply a MATLAB implementation of Algorithm 6.2 to several test problems. All computations are run under MATLAB 7.13 (R2011b) on an Intel® Core™ i7-2640M processor with 2.8 GHz and 4 GB of memory. The presented residuals are computed by

$$\frac{\|T(\theta)y + U(\theta)z\|_2}{\|T(\theta)\|_{\mathsf{F}}}$$

with U defined as in (6.3) and $\|\cdot\|_{\mathsf{F}}$ denoting the Frobenius norm.

Our implementation always solves the projected eigenproblems by the contour integral method, even for the last experiment, where the problem at hand is polynomial and hence eligible for linearization. The contours are circles around the current eigenvalue approximations, whose radius ρ depends on the problem. We employ the selective deflation strategy from Section 6.2.8 with safety factor $\gamma = 1.2$ and restart the algorithm as outlined in Section 6.2.7. For the deflation, we use the polynomial basis formed by the scaled and shifted monomials $p_i(\lambda) = \alpha^i \cdot (\lambda - \theta_0)^i$, $i = 0, 1, 2, \ldots$, where θ_0 signifies the current target and $\alpha = (\gamma\rho)^{-1}$.

The correction equations are solved iteratively by means of GMRES. The solver is stopped after at most 10 iterations or earlier if the residual has been decreased by a factor of 10^{-2}. The shift σ is updated in every step with the current eigenvalue approximation.

In every experiment, we start the Jacobi-Davidson algorithm using certain approximations to the desired eigenvalues. How these approximations are obtained is described within the individual subsections. The starting vectors, however, are always picked at random and we do not inject any prior knowledge about the eigenvectors.

6.3.1 Delay eigenvalue problem. As our first experiment, we consider the parabolic partial differential equation

$$\frac{\partial u}{\partial t}(x,t) = \frac{\partial^2 u}{\partial x^2}(x,t) + a_0 u(x,t) + a_1(x) u(x, t-\tau), \quad u(0,t) = u(\pi,t) = 0 \tag{6.28}$$

with time delay τ and coefficients $a_0 = 20$, $a_1(x) = -4.1 + x(1 - \mathrm{e}^{x-\pi})$, taken from [65, Sec. 2.4.1], which is a modification of [142, Chapter 3, Example 1.12]. Discretizing Equation (6.28) in space by means of finite differences on the uniform

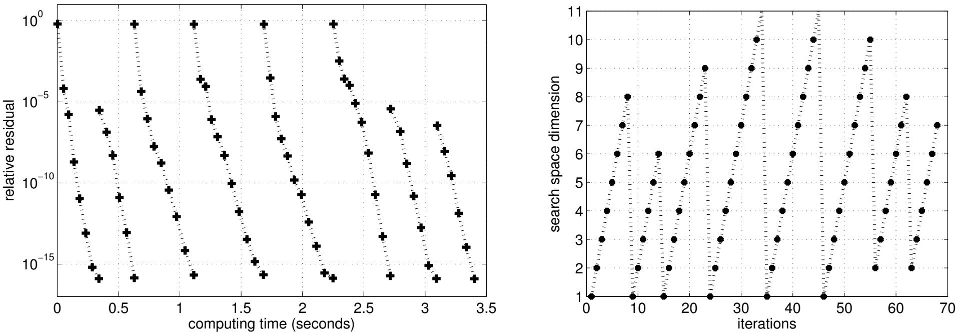

Figure 6.1: Left. Convergence history for the delay eigenvalue problem. Right. Evolution of the search space dimension.

grid $\{x_i = \frac{i}{n+1}\pi : i = 1, \dots, n\}$ of size $h = \frac{\pi}{n+1}$ leads to the delay differential equation

$$\dot{v}(t) = A_0 v(t) + A_1 v(t - \tau) \tag{6.29}$$

with $v(t) = \big[u(x_1, t), \dots, u(x_n, t)\big]^{\mathsf{T}}$ and the $n \times n$ coefficient matrices

$$A_0 = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & -2 \end{bmatrix} + a_0 I, \quad A_1 = \begin{bmatrix} a_1(x_1) & & \\ & \ddots & \\ & & a_1(x_n) \end{bmatrix}.$$

An asymptotic stability analysis of (6.29) requires a few eigenvalues with largest real part of the delay eigenvalue problem $T(\lambda)v = 0$ with

$$T(\lambda) = -\lambda I + A_0 + \mathrm{e}^{-\tau\lambda} A_1.$$

For our experiment, we choose $n = 1000$, $\tau = 0.2$. As a preconditioner for $T(\lambda)$, we employ the discrete second derivative

$$P = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & -2 \end{bmatrix}.$$

Note that $-P$ is positive definite and can thus be efficiently inverted via a sparse Cholesky decomposition. If λ has a large, positive real part, $\mathrm{e}^{-\tau\lambda}$ tends to be small, implying $T(\lambda) \approx A_0 - \lambda I$. This encourages using the eigenvalues of A_0 with largest real part as starting values.

The left plot in Figure 6.1 illustrates the convergence of the 8 largest real eigenvalues of T. The eigenvalues computed are in this order: 18.932251, 15.868175, 10.618574, 1.733673, -5.342532, -9.215977, -10.717667, -11.818305. The computations of the 2^{nd}, 7^{th}, and 8^{th} eigenvalue have been warm-started using the restarting technique from Section 6.2.7. As a result, their initial residuals are considerably lower than those of the other eigenvalues, where no information could

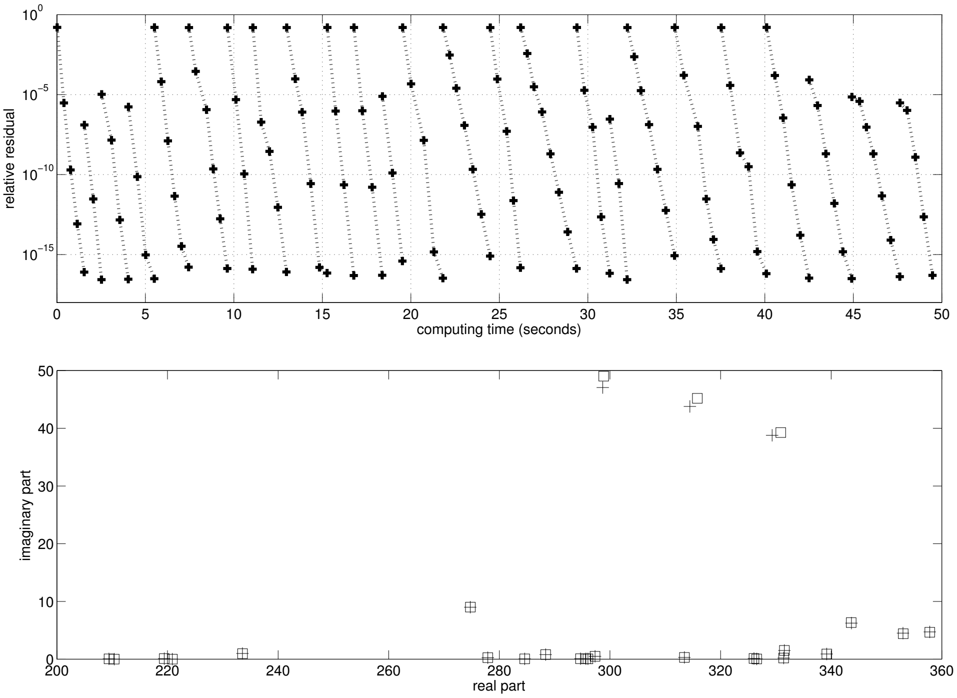

Figure 6.2: Top. Convergence history for the gun example. Bottom. Eigenvalues computed by Algorithm 6.2 ($+$) and approximate eigenvalues from linearized problem ($\square$).

be reused from the previous run. The evolution of the search space is depicted on the right in Figure 6.1; its dimension never exceeds 11.

6.3.2 Radio-frequency gun cavity. As a second example, we consider the nonlinear eigenvalue problem induced by the matrix-valued function

$$T(\lambda) = K - \lambda^2 M + \mathrm{i}\sqrt{\lambda^2 - \sigma_1^2}\, W_1 + \mathrm{i}\sqrt{\lambda^2 - \sigma_2^2}\, W_2,$$

taken from [15]. It models a radio-frequency gun cavity; for details, see also [88]. Here, $\sqrt{\cdot}$ denotes the principal branch of the complex square root. The matrices $K, M, W_1, W_2 \in \mathbb{R}^{9956\times 9956}$ are symmetric and sparse, $\sigma_1 = 0$, and $\sigma_2 = 108.8774$. Linearizing the square roots around $\lambda_0^2 - \sigma_i^2$, we find $T(\lambda) \approx A - \lambda^2 B$ with

$$\begin{aligned} A &= K + \frac{\mathrm{i}(\lambda_0^2 - 2\sigma_1^2)}{2\sqrt{\lambda_0^2 - \sigma_1^2}} W_1 + \frac{\mathrm{i}(\lambda_0^2 - 2\sigma_2^2)}{2\sqrt{\lambda_0^2 - \sigma_2^2}} W_2, \\ B &= M - \frac{\mathrm{i}}{2\sqrt{\lambda_0^2 - \sigma_1^2}} W_1 - \frac{\mathrm{i}}{2\sqrt{\lambda_0^2 - \sigma_2^2}} W_2, \end{aligned}$$

which is linear in λ^2 and provides excellent approximations for the eigenvalues close to λ_0.

For our experiment, we take $\lambda_0^2 = 81250$ and aim at computing the 25 eigenvalues of T closest to λ_0. The bottom part of Figure 6.2 displays the computed eigenvalues ($+$) along with the approximations ($\square$) from the linearized problem

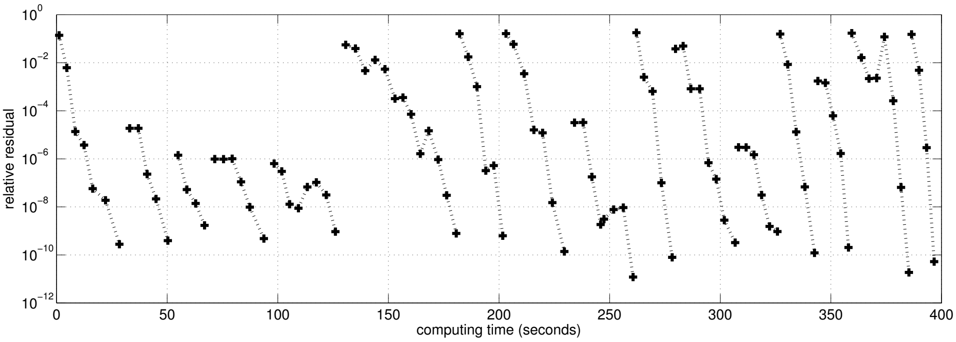

Figure 6.3: Convergence history for the boundary element example.

above. The abundance of tightly clustered eigenvalues in this example is particularly appealing for testing our algorithm, which exhibits no flaws in determining them correctly. The convergence history is presented in the top part of Figure 6.2. The merits of our restarting procedure are again clearly visible for eigenvalues number 2, 3, 4, 12, 18, 23, 24, 25.

6.3.3 Boundary-element discretization of Laplace eigenvalue problem. As our final example, we consider a polynomial interpolant

$$T(\lambda) = \tau_0(\lambda)P_0 + \cdots + \tau_d(\lambda)P_d$$

of a matrix-valued function stemming from a boundary-element discretization of the Laplace eigenvalue problem with Dirichlet boundary conditions on the unit cube; see Section 2.3 as well as Chapter 5. For stability reasons, the polynomial is given in the Chebyshev basis $\tau_0, \ldots, \tau_d$. The coefficient matrices $P_0, \ldots, P_d$ are complex of order 2400×2400.

We aim at computing the six smallest eigenvalues of T. In the continuous setting, the multiplicities of these eigenvalues are 1, 3, 3, 3, 1, 6. As a result, the polynomial T has very tight eigenvalue clusters, agreeing up to 10 decimal places, making it very hard for a Newton-based method. In consequence, we observe some non-monotonic convergence behavior in Figure 6.3. However, the algorithm does retrieve the correct number of eigenvalues in each cluster. Moreover, after the first eigenvalue in a cluster has converged, our restarting strategy helps to find the remaining ones more quickly. This is especially apparent from the six-fold eigenvalue, which has been computed first.

Contributions within this chapter

In this chapter, we have proposed a technique to deflate arbitrary minimal invariant pairs from a nonlinear eigenvalue problem, which needs to be given only as an oracle for the block residual. This technique greatly enhances the performance of Newton-based solvers by eliminating their tendency to reconverge to previously determined eigenpairs. The deflation takes place via a suitable bordering of the original nonlinear eigenvalue problem, leading to a problem of slightly increased size.

In Section 6.1, the bordered nonlinear eigenvalue problem is derived from a set of sufficient criteria for the expansion of an existing minimal invariant pair in Lemmas 6.1.1, 6.1.4, and 6.1.5. Afterwards, our central result in Theorem 6.1.6 demonstrates that this bordering indeed effects a deflation. All results of Section 6.1 have been published by the author in [39].

Section 6.2 incorporates this deflation technique into a Jacobi-Davidson-type algorithm. To this end, first a Newton-based method for solving the bordered nonlinear eigenvalue problems is developed in Section 6.2.1 and later subspace acceleration is added in Section 6.2.2. Subsequently, we discuss various algorithmic details required for a practical implementation of the algorithm. In Section 6.2.5, a preconditioner for the correction equation is constructed, based on a given preconditioner for the matrix-valued function, and it is shown that both preconditioners have roughly the same cost and quality. In Section 6.2.7, we develop a restarting strategy capable of handling the increase in problem size as additional eigenpairs are locked. Finally, we argue in Section 6.2.8 that only a limited number of eigenpairs are actually threatened by reconvergence and demonstrate how to selectively deflate only these eigenpairs. In conjunction with the custom preconditioner from Section 6.2.5, this measure helps to reduce the computational overhead incurred by the deflation to a minimum.

Chapter 7

Preconditioned solvers for nonlinear eigenvalue problems admitting a Rayleigh functional

In this chapter, we consider nonlinear eigenvalue problems for which a Rayleigh functional can be defined. Rayleigh functionals have been introduced in [38] for quadratic eigenvalue problems and extended to more general cases in [111, 55, 138]. They generalize the notion of the Rayleigh quotient for a Hermitian matrix to the nonlinear eigenvalue problem (1.1) where the matrix $T(\lambda)$ is Hermitian for every λ of interest, which will be assumed throughout this chapter.

We introduce the concept following [138]. Let $\mathcal{D}$ be a possibly unbounded, real interval and assume that for any non-zero vector $x \in \mathbb{C}^n$, the scalar function

$$f_x(\lambda) = x^{\mathsf{H}} T(\lambda) x$$

has at most one zero within the interval $\mathcal{D}$. This permits us to define a function $\rho : D(\rho) \to \mathcal{D}$ mapping a vector x onto the corresponding zero of f_x. Here, $D(\rho)$ is the set of all vectors $x \in \mathbb{C}^n \setminus \{0\}$ for which the zero exists. The function ρ is called a *Rayleigh functional* of T if for all $x \in D(\rho)$ and all $\lambda \in \mathcal{D}$, $\lambda \neq \rho(x)$,

$$\big(\lambda - \rho(x)\big) \cdot f_x(\lambda) > 0 \tag{7.1}$$

or, in other words, if the function f_x switches sign from negative to positive at $\rho(x)$. If T, and hence also f_x, is differentiable, condition (7.1) is equivalent to

$$\dot{f}_x\big(\rho(x)\big) > 0$$

for all $x \in D(\rho)$.

7.1 Perturbation theory for Rayleigh functionals

It has been shown in [138] that the domain of definition $D(\rho)$ of the Rayleigh functional ρ is an open set if the matrix-valued function T is continuous. Moreover, it is easily seen that for any eigenpair (x, λ) of the nonlinear eigenvalue problem, the eigenvector x is contained in $D(\rho)$ and $\rho(x) = \lambda$. Since ρ inherits the continuity of T, it follows that $\rho(y) \to \lambda$ as $y \to x$. If continuous differentiability of T is assumed, this convergence can be quantified. A corresponding result has appeared in [117, Corollary 18]. The subsequent proposition represents an improvement of this result in that it bounds the deviation of the Rayleigh functional from λ in terms of the sine of the angle between y and x instead of the tangent. Additionally, the proposition brings out more clearly that there is a trade-off between the constant in the bound (7.2) and the opening angle of the cone in which the bound holds. Lastly, the proof offered here is simpler than that in [117].

Proposition 7.1.1. *Suppose that the matrix-valued function T is differentiable and that its derivative $\dot{T}$ is Lipschitz continuous with Lipschitz constant $\gamma_{\dot{T}}$. Let $\|\cdot\|$ denote the Euclidean vector norm or some compatible matrix norm, depending on the context, and let (x, λ) with $\|x\| = 1$ be an eigenpair of T. Choose $0 < \theta < 1$ and define*

$$\delta = \theta \cdot \frac{x^{\mathsf{H}}\dot{T}(\lambda)x}{\beta\|\dot{T}(\lambda)\|}, \qquad \beta = 1 + \sqrt{1+\Phi}, \qquad \Phi = \frac{\theta}{1-\theta} \cdot \frac{\gamma_{\dot{T}}\|T(\lambda)\|}{2\|\dot{T}(\lambda)\|^2}.$$

Then every $y \neq 0$ with $\sin\angle(y, x) \leq \delta$ belongs to $D(\rho)$ and

$$|\rho(y) - \lambda| \leq \frac{1}{1-\theta} \cdot \frac{\|T(\lambda)\|}{x^{\mathsf{H}}\dot{T}(\lambda)x} \cdot \sin^2\angle(y, x). \tag{7.2}$$

Proof. Because neither $\angle(y, x)$ nor $\rho(y)$ depend on the scaling of y, we can rescale y such that $y = x - e$ with e orthogonal to y. In particular, this implies $\|y\|^2 + \|e\|^2 = \|x\|^2 = 1$ and $\|e\| = \sin\angle(y, x)$.

Since the function $f_y(\mu) := y^{\mathsf{H}}T(\mu)y$ inherits the smoothness of T, it holds that

$$f_y(\mu) = f_y(\lambda) + (\mu - \lambda) \cdot R(\mu), \qquad R(\mu) = \int_0^1 \dot{f}_y\big(\lambda + \tau(\mu - \lambda)\big)\,\mathrm{d}\tau. \tag{7.3}$$

Exploiting that x is an eigenvector, we find $f_y(\lambda) = (x - e)^{\mathsf{H}}T(\lambda)(x - e) = e^{\mathsf{H}}T(\lambda)e$ and conclude

$$|f_y(\lambda)| \leq \|T(\lambda)\|\|e\|^2 \leq \frac{\theta^2}{\beta^2} \cdot \frac{\|T(\lambda)\|}{\|\dot{T}(\lambda)\|^2} \cdot (x^{\mathsf{H}}\dot{T}(\lambda)x)^2. \tag{7.4}$$

Moreover, we have

$$\begin{aligned} R(\mu) &= \dot{f}_y(\lambda) + \int_0^1 \big[\dot{f}_y\big(\lambda + \tau(\mu - \lambda)\big) - \dot{f}_y(\lambda)\big]\,\mathrm{d}\tau \\ &\geq \dot{f}_y(\lambda) - \int_0^1 \big\|\dot{T}\big(\lambda + \tau(\mu - \lambda)\big) - \dot{T}(\lambda)\big\|\,\mathrm{d}\tau. \end{aligned}$$

Estimating both terms on the right-hand side individually, we find

$$\begin{aligned} \dot{f}_y(\lambda) &= x^{\mathsf{H}}\dot{T}(\lambda)x - e^{\mathsf{H}}\dot{T}(\lambda)x - y^{\mathsf{H}}\dot{T}(\lambda)e \geq x^{\mathsf{H}}\dot{T}(\lambda)x - 2\|\dot{T}(\lambda)\|\|e\| \\ &\geq (1 - \tfrac{2}{\beta}\theta)x^{\mathsf{H}}\dot{T}(\lambda)x \end{aligned}$$

and

$$\int_0^1 \big\|\dot{T}\big(\lambda + \tau(\mu - \lambda)\big) - \dot{T}(\lambda)\big\|\,\mathrm{d}\tau \leq \int_0^1 \gamma_{\dot{T}}\tau|\mu - \lambda|\,\mathrm{d}\tau = \tfrac{\gamma_{\dot{T}}}{2}|\mu - \lambda|.$$

For $\mu \in [\mu^-, \mu^+]$ with $\mu^\pm := \lambda \pm \frac{2}{\gamma_{\dot{T}}}(1-\frac{2}{\beta})\theta x^{\mathsf{H}}\dot{T}(\lambda)x$, combining the above estimates yields

$$R(\mu) \geq (1-\theta)x^{\mathsf{H}}\dot{T}(\lambda)x > 0.$$

By construction, β satisfies $\beta(\beta-2) = \Phi$, which is equivalent to $1 - \frac{2}{\beta} = \frac{\Phi}{\beta^2} > 0$. Therefore, using the definition of Φ and the estimate for $|f_y(\lambda)|$ in (7.4),

$$\begin{aligned} f_y(\mu^+) &\geq (\mu^+ - \lambda) \cdot R(\mu^+) - |f_y(\lambda)| \\ &\geq \frac{\theta(x^{\mathsf{H}}\dot{T}(\lambda)x)^2}{\beta^2}\left[\frac{2\Phi}{\gamma_{\dot{T}}}(1-\theta) - \theta \cdot \frac{\|T(\lambda)\|}{\|\dot{T}(\lambda)\|^2}\right] = 0. \end{aligned}$$

Analogously, one shows that $f_y(\mu^-) \leq 0$. Consequently, f_y has a zero inside the interval $[\mu^-, \mu^+]$, proving that $y \in D(\rho)$ and $\rho(y) \in [\mu^-, \mu^+]$, i.e., $|\rho(y) - \lambda| \leq \frac{2}{\gamma_{\dot{T}}}(1 - \frac{2}{\beta})\theta x^{\mathsf{H}}\dot{T}(\lambda)x$. Inserting $\mu = \rho(y)$ into the expansion (7.3) and rearranging leads to the bound

$$|\rho(y) - \lambda| = \frac{|f_y(\lambda)|}{|R(\rho(y))|} \leq \frac{\|T(\lambda)\|\|e\|^2}{(1-\theta)x^{\mathsf{H}}\dot{T}(\lambda)x}$$

as claimed. □

7.2 Preconditioned residual inverse iteration

The existence of a Rayleigh functional ρ has a number of fundamental consequences for the underlying nonlinear eigenvalue problem. In particular, a Rayleigh functional enables the characterization of the eigenvalues within the interval $\mathcal{D}$ by means of three variational principles; see [56, 139, 132] as well as Section 1.2.5. Specifically, if

$$\lambda_1 := \inf_{x \in D(\rho)} \rho(x)$$

is contained in the interval $\mathcal{D}$, then λ_1 is the first eigenvalue of T in $\mathcal{D}$, and the infimum is attained for the corresponding eigenvector x_1. This fact suggests that the eigenpair (x_1, λ_1) can be computed through Rayleigh functional minimization. In the case of a linear eigenvalue problem for a Hermitian matrix $A \in \mathbb{C}^{n\times n}$, $T(\lambda) = \lambda I - A$, this strategy leads to the preconditioned inverse iteration [102, 103] for determining the smallest eigenvalue, given by

$$v_{j+1} = v_j + P^{-1}\left(\frac{v_j^{\mathsf{H}}Av_j}{v_j^{\mathsf{H}}v_j}I - A\right)v_j. \tag{7.5}$$

Here, P is a positive definite preconditioner for A. Generalizing this iteration to the nonlinear case yields

$$v_{j+1} = v_j + P^{-1}T\big(\rho(v_j)\big)v_j. \tag{7.6}$$

Remark 7.2.1. The iteration (7.6) can also be viewed as one step of safeguarded iteration (see Section 1.2.5), where the exact solution of the auxiliary linear eigenvalue problem is replaced by one step of the preconditioned inverse iteration (7.5). More precisely, let v_j be the current iterate. Then the next iterate produced by the safeguarded iteration would be the eigenvector corresponding to the largest eigenvalue of $T\big(\rho(v_j)\big)$ or, alternatively, the eigenvector belonging to the smallest eigenvalue of $-T\big(\rho(v_j)\big)$. One step of the preconditioned inverse iteration (7.5) with starting vector v_j applied to the latter problem reads

$$v_{j+1} = v_j + P^{-1}\bigg(-\frac{v_j^{\mathsf{H}} T\big(\rho(v_j)\big)v_j}{v_j^{\mathsf{H}} v_j} I + T\big(\rho(v_j)\big)\bigg)v_j.$$

Since, by definition of the Rayleigh functional, $v_j^{\mathsf{H}} T\big(\rho(v_j)\big)v_j = 0$, this is exactly the nonlinear iteration (7.6).

Let $\sigma \in \mathcal{D}$ with $\sigma < \lambda_1$. Then, $T(\sigma)$ is negative definite by [82, Theorem 3.2 (i)]. If $P = -T(\sigma)$ is employed as the preconditioner, the nonlinear iteration (7.6) becomes identical to the residual inverse iteration derived in [101]. The residual inverse iteration has been shown in [101, Section 3] to converge linearly towards eigenvectors associated with simple eigenvalues, with a convergence factor roughly proportional to the distance between σ and the eigenvalue λ_1. A further analysis of the convergence factor for the residual inverse iteration has been conducted in [68].

In the following, we will analyze the convergence of the iteration (7.6) for an arbitrary Hermitian positive definite preconditioner P. For the analysis, it will be most convenient to work in the geometry induced by the P-inner product, $\langle\cdot,\cdot\rangle_P$. In particular, we will denote the angle between two vectors in the P-inner product by $\angle_P(\cdot,\cdot)$, as opposed to $\angle_2(\cdot,\cdot)$, which measures the angle in the Euclidean inner product. Moreover, we will write $\|\cdot\|_P$ for the norm induced by P.

Theorem 7.2.2. *Let the matrix-valued function T be differentiable with Lipschitz continuous derivative, and let λ_1 be the first eigenvalue of T in $\mathcal{D}$ with corresponding eigenvector x_1. Suppose that the preconditioner P is Hermitian positive definite and spectrally equivalent to $-T(\lambda_1)$ on the P-orthogonal complement of the eigenvector x_1, i.e., there is some $\gamma < 1$ such that*

$$(1-\gamma)y^{\mathsf{H}}Py \le -y^{\mathsf{H}}T(\lambda_1)y \le (1+\gamma)y^{\mathsf{H}}Py \qquad \forall y \neq 0,\ \langle y, x_1\rangle_P = 0. \tag{7.7}$$

Then one step of the preconditioned residual inverse iteration (7.6) *satisfies*

$$\tan\angle_P(v_{j+1}, x_1) \le \gamma\epsilon_j + O(\epsilon_j^2),$$

provided that $\epsilon_j := \tan\angle_P(v_j, x_1)$ is sufficiently small.

Proof. Define $T_P : \mathcal{D} \to \mathbb{C}^{n\times n}$ via $T_P(\lambda) := P^{-\frac{1}{2}}T(\lambda)P^{-\frac{1}{2}}$. One then readily verifies that T_P is differentiable with a Lipschitz continuous derivative. Moreover, it is easy to see that $\rho_P(x) := \rho(P^{-\frac{1}{2}}x)$ defines a Rayleigh functional for T_P and $(P^{\frac{1}{2}}x_1, \lambda_1)$ constitutes an eigenpair of T_P. Let $\angle_P(v_j, x_1) = \angle_2(P^{\frac{1}{2}}v_j, P^{\frac{1}{2}}x_1)$ be sufficiently small. Then, by Proposition 7.1.1, there exists a constant $C > 0$ for which

$$|\rho(v_j) - \lambda_1| = |\rho_P(P^{\frac{1}{2}}v_j) - \lambda_1| \le C\cdot\sin^2\angle_2(P^{\frac{1}{2}}v_j, P^{\frac{1}{2}}x_1) = C\cdot\sin^2\angle_P(v_j, x_1).$$

Exploiting that $\sin\angle_P(v_j, x_1) \le \tan\angle_P(v_j, x_1)$, we therefore obtain

$$\rho(v_j) = \lambda_1 + O(\epsilon_j^2).$$

Together, with a Taylor expansion of T,

$$T\big(\rho(v_j)\big) = T(\lambda_1) + \big(\rho(v_j) - \lambda_1\big)\cdot\dot{T}(\lambda_1) + o(|\rho(v_j) - \lambda_1|),$$

this implies

$$T\big(\rho(v_j)\big) = T(\lambda_1) + O(\epsilon_j^2). \tag{7.8}$$

Decomposing $v_j = v_j^{(\|)} + v_j^{(\perp)}$ with $v_j^{(\|)} \in \operatorname{span}\{x_1\}$ and $\big\langle v_j^{(\perp)}, v_j^{(\|)}\big\rangle_P = 0$, the next iterate reads

$$v_{j+1} = v_j^{(\|)} + v_j^{(\perp)} + P^{-1}T\big(\rho(v_j)\big)\big(v_j^{(\|)} + v_j^{(\perp)}\big).$$

Employing (7.8) and taking into account that $T(\lambda_1)v_j^{(\|)} = 0$, this becomes

$$v_{j+1} = v_j^{(\|)} + v_j^{(\perp)} + P^{-1}T(\lambda_1)v_j^{(\perp)} + O(\epsilon_j^2).$$

Observing that $\big\langle P^{-1}T(\lambda_1)v_j^{(\perp)}, x_1\big\rangle_P = 0$, a corresponding decomposition of v_{j+1} is given by $v_{j+1} = v_{j+1}^{(\|)} + v_{j+1}^{(\perp)}$ with

$$v_{j+1}^{(\|)} = v_j^{(\|)} + O(\epsilon_j^2), \qquad v_{j+1}^{(\perp)} = v_j^{(\perp)} + P^{-1}T(\lambda_1)v_j^{(\perp)} + O(\epsilon_j^2).$$

Consequently,

$$\begin{aligned} \tan\angle_P(v_{j+1}, x_1) &= \frac{\big\|v_{j+1}^{(\perp)}\big\|_P}{\big\|v_{j+1}^{(\|)}\big\|_P} = \frac{\big\|\big(I + P^{-1}T(\lambda_1)\big)v_j^{(\perp)}\big\|_P}{\big\|v_j^{(\|)}\big\|_P} + O(\epsilon_j^2) \\ &\leq \sup_{\substack{y \neq 0 \\ \langle y, x_1\rangle_P = 0}} \frac{\big\|\big(I + P^{-1}T(\lambda_1)\big)y\big\|_P}{\|y\|_P} \cdot \tan\angle_P(v_j, x_1) + O(\epsilon_j^2). \end{aligned}$$

Setting $P^{\frac{1}{2}}y =: z$, one calculates

$$\sup_{\substack{y \neq 0 \\ \langle y, x_1\rangle_P = 0}} \frac{\big\|\big(I + P^{-1}T(\lambda_1)\big)y\big\|_P}{\|y\|_P} = \sup_{\substack{z \neq 0 \\ \langle z, P^{1/2}x_1\rangle = 0}} \frac{\big\|\big(I + P^{-\frac{1}{2}}T(\lambda_1)P^{-\frac{1}{2}}\big)z\big\|_2}{\|z\|_2}.$$

Because $P^{\frac{1}{2}}x_1$ is an eigenvector of the Hermitian matrix $I + P^{-\frac{1}{2}}T(\lambda_1)P^{-\frac{1}{2}}$, the supremum on the right equals

$$\sup_{\substack{z \neq 0 \\ \langle z, P^{1/2}x_1\rangle = 0}} \left|\frac{z^{\mathsf{H}}\big(I + P^{-\frac{1}{2}}T(\lambda_1)P^{-\frac{1}{2}}\big)z}{z^{\mathsf{H}}z}\right| = \sup_{\substack{y \neq 0 \\ \langle y, x_1\rangle_P = 0}} \left|1 + \frac{y^{\mathsf{H}}T(\lambda_1)y}{y^{\mathsf{H}}Py}\right|,$$

which, by the spectral equivalence (7.7), cannot exceed γ. □

In short, Theorem 7.2.2 states that the preconditioned residual inverse iteration (7.6) converges linearly towards the first eigenvalue λ_1 if the preconditioner utilized is spectrally equivalent to $-T(\lambda_1)$ on the P-orthogonal complement of the corresponding eigenvector x_1. It has been shown in [68, Theorem 4.2] that such a preconditioner is given by $P = -T(\sigma)$ if $0 < \lambda_1 - \sigma$ is sufficiently small. However, in practice, linear systems with the matrix $-T(\sigma)$ might be expensive to solve. In this event, one may wish to resort to a preconditioner which is easier to apply and spectrally equivalent to $-T(\sigma)$ instead of employing $-T(\sigma)$ itself.

We even conjecture that mesh-independent convergence rates can be achieved for a sequence of matrix-valued functions T_k, $k = 1, 2, \dots$ representing increasingly fine discretizations of a nonlinear operator eigenvalue problem by selecting preconditioners P_k, $k = 1, 2, \dots$ which satisfy

$$(1-\gamma)y^{\mathsf{H}}P_k \leq -y^{\mathsf{H}}T_k(\sigma)y^{\mathsf{H}} \leq (1+\gamma)y^{\mathsf{H}}P_k y \qquad \forall y \neq 0$$

with a uniform constant $\gamma < 1$ for all $k = 1, 2, \dots$. Examples for preconditioners having this property include the W-cycle and generalized V-cycle symmetric multigrid methods analyzed in [26] as well as the BPX preconditioner [27]. Although mesh-independent convergence is observed in the numerical experiments of the next section, a rigorous mathematical proof remains to be done.

7.3 Numerical experiments

As an example, we consider the vibrating string with an elastically attached mass from Section 2.4. As established there, this problem admits a Rayleigh functional for the interval of interest. We aim at computing the first eigenvalue within this interval, which, for the continuous problem, amounts to $\lambda_1 = 4.482024295$.

Discretizing the problem by linear finite elements on a uniform grid of size h leads to a nonlinear matrix eigenvalue problem of the form (2.16). The corresponding matrix-valued function will be abbreviated by T_h. We solve this problem by the preconditioned residual inverse iteration (7.6) for different levels of mesh refinement. As preconditioner, we employ one W-cycle of the symmetric multigrid algorithm with Jacobi smoother described in [53], applied to $-T_h(\sigma)$ for a suitably chosen $\sigma < \lambda_1$. It has been shown in [26] that this preconditioner is spectrally equivalent to $-T_h(\sigma)$, uniformly with respect to the grid size h.

	#iterations		
h	$\sigma = 0$	$\sigma = 2$	$\sigma = 4$
2^{-5}	12	8	5
2^{-6}	12	8	5
2^{-7}	12	8	5
2^{-8}	12	7	4
2^{-9}	12	7	4
2^{-10}	12	7	4
2^{-11}	11	7	4
2^{-12}	11	7	4
2^{-13}	11	6	4
2^{-14}	10	6	4
2^{-15}	10	6	4
2^{-16}	10	6	4
2^{-17}	9	6	3
2^{-18}	9	5	3

Table 7.1: Iteration numbers of the preconditioned residual inverse iteration (7.6) applied to the vibrating string problem for different grid sizes h. The employed preconditioners are spectrally equivalent to $-T_h(\sigma)$ for different values of σ.

Table 7.1 reports the numbers of iterations required to reach an accuracy of 10^{-6} in the eigenvalue for different levels of mesh refinement and different choices of the parameter σ. Clearly, the iteration numbers do not increase as the mesh is further refined; in fact, they even slightly decrease. This lends numerical evidence to our conjecture that the use of uniformly spectrally equivalent preconditioners results in mesh-independent convergence. Moreover, we observe that the number of iterations reduces as σ moves closer to λ_1 as expected. All computations have been performed under MATLAB 7.13 (R2011b).

Contributions within this chapter

In this chapter, we have considered nonlinear eigenvalue problems for which a Rayleigh functional can be defined. After introducing the concept, we derive a preconditioned nonlinear eigensolver based on Rayleigh functional minimization in analogy to corresponding techniques for linear, Hermitian eigenvalue problems. We conjecture that, as in the linear case, this method leads to mesh-independent convergence rates for discretizations of operator eigenvalue problems when using suitable preconditioners.

In Section 7.1, we study the behavior of the Rayleigh functional for slightly perturbed eigenvectors. Proposition 7.1.1 provides a corresponding bound, which generalizes an existing result in [117, Corollary 18]. Furthermore, the proof given here is simpler than the original one.

In Section 7.2, we generalize the preconditioned inverse iteration [102, 103] for linear, Hermitian eigenvalue problems to the nonlinear setting. For a specific choice of the preconditioner, the resulting method coincides with the residual inverse iteration [101]. Furthermore, we demonstrate in Remark 7.2.1 that the method can also be viewed as an inexact version of the safeguarded iteration proposed in [140]. Theorem 7.2.2 analyzes the convergence of our method, extending and improving an existing analysis of the residual inverse iteration in [68].

Section 7.3 provides numerical evidence that the derived method can deliver mesh-independent convergence if an appropriate preconditioner is employed. A formal mathematical proof of this fact, however, remains open.

Chapter 8

Conclusion

In Chapter 3, we have compiled and substantially extended the existing theory on minimal invariant pairs, laying the foundations for the developments in the remainder of the work. In particular, we have given a broader definition of invariance, which does not require a specific form of the matrix-valued function. Furthermore, we have provided a more versatile characterization of minimality by allowing more freedom in the selection of the underlying polynomial basis. In both cases, we have established the equivalence of the new formulations with the original ones under certain natural assumptions.

We have related the block residuals of composite pairs to certain derivatives of the block residual for their constituents. Based on these relationship, we have argued that an oracle for the block residual represents an excellent user-interface for virtually all nonlinear eigensolvers, which is far more flexible than the interfaces currently in use. It would be worthwhile to strengthen this argument by redesigning the existing nonlinear eigensolvers to utilize this new interface.

Finally, we have enhanced an existing theorem about the extraction of minimal invariant pairs from non-minimal ones. Furthermore, this technique is complemented with a result on the embedding of minimal invariant pairs into simple ones. We have shown that a combination of these two strategies can lead to very elegant and simple proofs, which we have exploited to prove the main results in Chapter 4.

In Chapter 4, we have developed a numerical scheme for simultaneously continuing several eigenvalues and (generalized) eigenvectors of a nonlinear eigenvalue problem, extending previous work in [22] for linear and [23] for quadratic eigenvalue problems. In this context, the concept of minimal invariant pairs has proved an adequate nonlinear substitute for invariant subspaces, which have been successfully used for continuation in the linear case.

Our approach is based on characterizing a minimal invariant pair representing the eigenpairs of interest as a zero of a certain nonlinear function. The continuation of the minimal invariant pair then amounts to continuing the zero. A new, direct, and simpler proof is given for the fact that the Jacobian of the nonlinear function is nonsingular if and only if the continued minimal invariant pair is simple. Moreover, we completely characterize the null space of the Jacobian for non-simple invariant pairs, demonstrating a one-to-one correspondence between the elements of the null space and possible extensions of the minimal invariant pair. Using these results, we characterize the generic bifurcations that might be encountered during the continuation process and show how to treat them numerically by enlarging the continued minimal invariant pair.

It should be mentioned that the bordered Bartels-Stewart algorithm utilized to solve the arising linear systems has quadratic complexity with respect to the number of eigenpairs to be continued. Hence, our implementation is not well suited for large-scale problems. It seems that this complexity can be brought down to linear by using specialized techniques from [50]. However, the details remain to be investigated.

The selection of the minimality index ℓ in the normalization condition (4.2) offers further room for improvement. In our implementation, ℓ is fixed throughout the entire procedure. It would, however, be desirable to choose ℓ adaptively in every step. This would potentially save some work by preventing overestimation of the minimality index.

In Chapter 5, we have presented a new approach to the approximate solution of nonlinear eigenvalue problems, based on approximating the nonlinear matrix-valued function by a suitable interpolating polynomial. This approach has several advantages. First of all, the construction of the polynomial interpolant requires only a few evaluations of the matrix-valued function at a set of interpolation nodes in the region of interest. This feature is especially attractive for nonlinear eigenvalue problems where evaluating the matrix-valued function is very expensive. Boundary-element discretizations of operator eigenvalue problems typically lead to such problems. The method thus overcomes a major bottleneck in the application of boundary-element discretizations to PDE eigenvalue problems, promoting the use of boundary elements in this area. Furthermore, the method does not require any derivatives of the matrix-valued function, which may not be readily available depending on the application.

Since the interpolant is a polynomial, its eigensystem can be conveniently computed through linearization, i.e., conversion into a linear eigenvalue problem of larger size—an option which is not available for the original nonlinear eigenvalue problem. We have shown how the linearized problem can be solved efficiently by means of Krylov subspace methods without the need to set up the defining matrices explicitly.

In a rigorous error analysis, it has been demonstrated that (part of) the eigensystem of the interpolating polynomial provides good approximations to the eigenvalues of the original problem in the region of interest as well as their associated eigenvectors. The first-order perturbation result for simple invariant pairs, which has been derived as part of this analysis, may as well be of independent interest.

It should, however, be mentioned that due to the use of Chebyshev interpolation, the method is confined to problems where to eigenvalues of interest lie on or close to a real interval. It would be interesting to consider also other interpolation schemes which are not subject to this restriction.

The method has been successfully tested in the context of boundary-element discretizations of PDE eigenvalue problems; its potential for other applications remains to be explored. To address more challenging problems, the implementation should be improved to take advantage of the techniques discussed in Remark 5.3.1.

In Chapter 6, we have presented a method to deflate arbitrary minimal invariant pairs from a given nonlinear eigenvalue problem. This technique greatly enhances the performance of Newton-based methods by eliminating their tendency to reconverge to previously determined eigenpairs. The deflation takes place via a suitable bordering of the original nonlinear eigenvalue problem, leading to a problem of slightly increased size.

We have incorporated this deflation strategy into a Jacobi-Davidson-type algorithm. Various algorithmic details have been discussed, including a restarting strategy capable of handling the increase in problem size as additional eigenpairs are locked. We have demonstrated that by employing a suitable preconditioner for the correction equation and selectively deflating only specific eigenpairs, the computational overhead incurred by the deflation is kept to a minimum.

The effectiveness of the ensuing algorithm has been demonstrated by means of several numerical experiments. Moreover, it should be emphasized that the algorithm only requires the block residual and a preconditioner for the original nonlinear eigenvalue problem to run. Thus, it is qualified for implementation as a general-purpose solver.

In Chapter 7, we have considered a preconditioned version of the residual inverse iteration for nonlinear eigenvalue problems admitting a Rayleigh functional. Alternatively, this method can be viewed as an inexact form of the safeguarded iteration. We have analyzed the convergence of the proposed method, extending previous work on the convergence of the residual inverse iteration in [68]. Moreover, we have conjectured that mesh-independent convergence can be achieved for discretizations of nonlinear operator eigenvalue problems by using appropriate preconditioners. This conjecture has been confirmed by a numerical experiment; however, a formal mathematical proof remains open. Before attempting such a proof, a precise mathematical framework for the infinite-dimensional problems needs to be defined, which currently requires more research.

Bibliography

[1] E. L. Allgower and K. Georg. *Numerical Continuation Methods*, volume 13 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1990. An introduction.

[2] T. Amako, Y. Yamamoto, and S. Zhang. A large-grained parallel algorithm for nonlinear eigenvalue problems and its implementation using OmniRPC. In *Cluster Computing, 2008 IEEE International Conference on*, pages 42–49, 2008.

[3] A. Amiraslani, R. M. Corless, and P. Lancaster. Linearization of matrix polynomials expressed in polynomial bases. *IMA J. Numer. Anal.*, 29(1):141–157, 2009.

[4] P. M. Anselone and L. B. Rall. The solution of characteristic value-vector problems by Newton's method. *Numer. Math.*, 11:38–45, 1968.

[5] V. I. Arnol'd. Matrices depending on parameters. *Uspehi Mat. Nauk*, 26(2(158)):101–114, 1971.

[6] V. I. Arnol'd. Lectures on bifurcations and versal families. *Uspehi Mat. Nauk*, 27(5(167)):119–184, 1972. A series of articles on the theory of singularities of smooth mappings.

[7] J. Asakura, T. Sakurai, H. Tadano, T. Ikegami, and K. Kimura. A numerical method for nonlinear eigenvalue problems using contour integrals. *JSIAM Letters*, 1:52–55, 2009.

[8] F. V. Atkinson. A spectral problem for completely continuous operators. *Acta Math. Acad. Sci. Hungar.*, 3:53–60, 1952.

[9] W. Axmann and P. Kuchment. An efficient finite element method for computing spectra of photonic and acoustic band-gap materials: I. Scalar case. *Journal of Computational Physics*, 150(2):468–481, 1999.

[10] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, editors. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, volume 11 of *Software, Environments, and Tools*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[11] C. G. Baker, U. L. Hetmaniuk, R. B. Lehoucq, and H. K. Thornquist. Anasazi software for the numerical solution of large-scale eigenvalue problems. *ACM Trans. Math. Software*, 36(3):Art. 13, 23, 2009.

[12] Z. Battles and L. N. Trefethen. An extension of MATLAB to continuous functions and operators. *SIAM J. Sci. Comput.*, 25(5):1743–1770, 2004.

[13] M. Bebendorf. Hierarchical LU decomposition-based preconditioners for BEM. *Computing*, 74(3):225–247, 2005.

[14] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numer.*, 14:1–137, 2005.

[15] T. Betcke, N. J. Higham, V. Mehrmann, C. Schröder, and F. Tisseur. NLEVP: A collection of nonlinear eigenvalue problems. MIMS EPrint 2011.116, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Dec. 2011.

[16] T. Betcke and D. Kressner. Perturbation, extraction and refinement of invariant pairs for matrix polynomials. *Linear Algebra Appl.*, 435(3):574–536, 2011.

[17] T. Betcke and L. N. Trefethen. Reviving the method of particular solutions. *SIAM Rev.*, 47(3):469–491, 2005.

[18] T. Betcke and H. Voss. A Jacobi-Davidson-type projection method for nonlinear eigenvalue problems. *Future Generation Computer Systems*, 20(3):363–372, 2004.

[19] W.-J. Beyn. An integral method for solving nonlinear eigenvalue problems. *Linear Algebra Appl.*, 436(10):3839–3863, 2012.

[20] W.-J. Beyn, A. Champneys, E. Doedel, W. Govaerts, Y. A. Kuznetsov, and B. Sandstede. Numerical continuation, and computation of normal forms. In *Handbook of dynamical systems, Vol. 2*, pages 149–219. North-Holland, Amsterdam, 2002.

[21] W.-J. Beyn, C. Effenberger, and D. Kressner. Continuation of eigenvalues and invariant pairs for parameterized nonlinear eigenvalue problems. *Numer. Math.*, 119(3):489–516, 2011.

[22] W.-J. Beyn, W. Kleß, and V. Thümmler. Continuation of low-dimensional invariant subspaces in dynamical systems of large dimension. In *Ergodic theory, analysis, and efficient simulation of dynamical systems*, pages 47–72. Springer, Berlin, 2001.

[23] W.-J. Beyn and V. Thümmler. Continuation of invariant subspaces for parameterized quadratic eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 31(3):1361–1381, 2009.

[24] D. Bindel, J. Demmel, and M. Friedman. Continuation of invariant subspaces in large bifurcation problems. *SIAM J. Sci. Comput.*, 30(2):637–656, 2008.

[25] M. A. Botchev, G. L. G. Sleijpen, and A. Sopaheluwakan. An SVD-approach to Jacobi-Davidson solution of nonlinear Helmholtz eigenvalue problems. *Linear Algebra Appl.*, 431(3-4):427–440, 2009.

[26] J. H. Bramble and J. E. Pasciak. New convergence estimates for multigrid algorithms. *Math. Comp.*, 49(180):311–329, 1987.

[27] J. H. Bramble, J. E. Pasciak, and J. Xu. Parallel multilevel preconditioners. *Math. Comp.*, 55(191):1–22, 1990.

[28] M. Cessenat. *Mathematical methods in electromagnetism*, volume 41 of *Series on Advances in Mathematics for Applied Sciences*. World Scientific Publishing Co. Inc., River Edge, NJ, 1996. Linear theory and applications.

[29] S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM J. Sci. Comput.*, 32(5):2737–2764, 2010.

[30] J.T. Chen, C.X. Huang, and K.H. Chen. Determination of spurious eigenvalues and multiplicities of true eigenvalues using the real-part dual BEM. *Computational Mechanics*, 24(1):41–51, 1999.

[31] J. Coomer, M. Lazarus, R.W. Tucker, D. Kershaw, and A. Tegman. A nonlinear eigenvalue problem associated with inextensible whirling strings. *Journal of Sound and Vibration*, 239(5):969 – 982, 2001.

[32] P. I. Davies and N. J. Higham. A Schur-Parlett algorithm for computing matrix functions. *SIAM J. Matrix Anal. Appl.*, 25(2):464–485, 2003.

[33] P. J. Davis. *Interpolation and Approximation*. Blaisdell Publishing Co., Ginn and Co., New York-Toronto-London, 1963.

[34] J. E. Dennis, Jr., J. F. Traub, and R. P. Weber. Algorithms for solvents of matrix polynomials. *SIAM J. Numer. Anal.*, 15(3):523–533, 1978.

[35] P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*, volume 35 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, corr. 2nd printing edition edition, 2006.

[36] L. Dieci and M. J. Friedman. Continuation of invariant subspaces. *Numer. Linear Algebra Appl.*, 8(5):317–327, 2001.

[37] D. C. Dobson. An efficient method for band structure calculations in 2D photonic crystals. *Journal of Computational Physics*, 149(2):363–376, 1999.

[38] R. J. Duffin. A minimax theory for overdamped networks. *J. Rational Mech. Anal.*, 4:221–233, 1955.

[39] C. Effenberger. Robust successive computation of eigenpairs for nonlinear eigenvalue problems. MATHICSE Technical Report 27.2012, EPF Lausanne, 2012.

[40] C. Effenberger and D. Kressner. Chebyshev interpolation for nonlinear eigenvalue problems. *BIT*, 52(4):933–951, 2012.

[41] C. Effenberger, D. Kressner, O. Steinbach, and G. Unger. Interpolation-based solution of a nonlinear eigenvalue problem in fluid-structure interaction. *PAMM*, 12(1):633–634, 2012.

[42] C. Engström. On the spectrum of a holomorphic operator-valued function with applications to absorptive photonic crystals. *Math. Models Methods Appl. Sci.*, 20(8):1319–1341, 2010.

[43] C. Engström and M. Wang. Complex dispersion relation calculations with the symmetric interior penalty method. *Internat. J. Numer. Methods Engrg.*, 84(7):849–863, 2010.

[44] A. Figotin and Y. A. Godin. The computation of spectra of some 2D photonic crystals. *Journal of Computational Physics*, 136(2):585–598, 1997.

[45] A. Figotin and P. Kuchment. Band-gap structure of spectra of periodic dielectric and acoustic media. I. Scalar model. *SIAM J. Appl. Math.*, 56(1):68–88, 1996.

[46] D. R. Fokkema, G. L. G. Sleijpen, and H. A. Van der Vorst. Jacobi-Davidson style QR and QZ algorithms for the reduction of matrix pencils. *SIAM J. Sci. Comput.*, 20(1):94–125, 1998.

[47] I. Gohberg, M. A. Kaashoek, and F. van Schagen. On the local theory of regular analytic matrix functions. *Linear Algebra Appl.*, 182:9–25, 1993.

[48] I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1982. Computer Science and Applied Mathematics.

[49] I. C. Gohberg and E. I. Sigal. An operator generalization of the logarithmic residue theorem and Rouché's theorem. *Math. USSR Sbornik*, 13(4):603–625, 1971.

[50] W. J. F. Govaerts. *Numerical Methods for Bifurcations of Dynamical Equilibria*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[51] J.-S. Guo, W. W. Lin, and C. S. Wang. Nonequivalence deflation for the solution of matrix latent value problems. *Linear Algebra Appl.*, 231:15–45, 1995.

[52] J.-S. Guo, W. W. Lin, and C. S. Wang. Numerical solutions for large sparse quadratic eigenvalue problems. *Linear Algebra Appl.*, 225:57–89, 1995.

[53] W. Hackbusch. *Multi-Grid Methods and Applications*. Number 4 in Springer Series in Computational Mathematics. Springer, Berlin, 1985.

[54] W. Hackbusch, L. Grasedyck, and S. Börm. An introduction to hierarchical matrices. In *Proceedings of EQUADIFF, 10 (Prague, 2001)*, volume 127, pages 229–241, 2002.

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

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

[57] M. L. J. Hautus. Controllability and observability conditions of linear autonomous systems. *Nederl. Akad. Wetensch. Proc. Ser. A 72 = Indag. Math.*, 31:443–448, 1969.

[58] M. Hein and T. Bühler. An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 847–855. 2010.

[59] V. Hernandez, J. E. Roman, and V. Vidal. SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Trans. Math. Software*, 31(3):351–362, 2005.

[60] N. J. Higham. *Functions of Matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008. Theory and computation.

[61] K. C. Huang, E. Lidorikis, X. Jiang, J. D. Joannopoulos, K. A. Nelson, P. Bienstman, and S. Fan. Nature of lossy bloch states in polaritonic photonic crystals. *Phys. Rev. B*, 69:195111, 2004.

[62] F. Ihlenburg. *Finite Element Analysis of Acoustic Scattering*, volume 132 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1998.

[63] I. C. F. Ipsen. A note on preconditioning nonsymmetric matrices. *SIAM J. Sci. Comput.*, 23(3):1050–1051, 2001.

[64] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons Inc., New York, third edition, 1999.

[65] E. Jarlebring. *The spectrum of delay-differential equations: numerical methods, stability and perturbation*. PhD thesis, Inst. Comp. Math, TU Braunschweig, 2008.

[66] E. Jarlebring and S. Güttel. A spatially adaptive iterative method for a class of nonlinear operator eigenproblems. MIMS EPrint 2012.113, Manchester Institute for Mathematical Sciences, School of Mathematics, The University of Manchester, Manchester, UK, November 2012.

[67] E. Jarlebring, K. Meerbergen, and W. Michiels. Computing a partial Schur factorization of nonlinear eigenvalue problems using the infinite Arnoldi method. Technical report, Dept. Comp. Science, K.U. Leuven, 2011.

[68] E. Jarlebring and W. Michiels. Analyzing the convergence factor of residual inverse iteration. *BIT*, 51(4):937–957, 2011.

[69] E. Jarlebring, W. Michiels, and K. Meerbergen. A linear eigenvalue algorithm for the nonlinear eigenvalue problem. *Numer. Math.*, 122(1):169–195, 2012.

[70] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade. *Photonic Crystals: Molding the Flow of Light*. Princeton University Press, Princeton, NJ, USA, second edition, 2008.

[71] T. Kailath. *Linear Systems*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1980. Prentice-Hall Information and System Sciences Series.

[72] N. Kamiya, E. Andoh, and K. Nogae. Eigenvalue analysis by the boundary element method: new developments. *Engineering Analysis with Boundary Elements*, 12(3):151–162, 1993.

[73] N. Kamiya and S. T. Wu. Generalized eigenvalue formulation of the Helmholtz equation by the Trefftz method. *Engineering Computations*, 11:177–186, 1994.

[74] M. Karow, D. Kressner, and E. Mengi. Nonlinear eigenvalue problems with specified eigenvalues. MATHICSE Technical Report, EPF Lausanne, 2013.

[75] M. V. Keldysh. On the completeness of the eigenfunctions of some classes of non-selfadjoint linear operators. *Russian Mathematical Surveys*, 26(4):15–44, 1971.

[76] H. B. Keller. Numerical solution of bifurcation and nonlinear eigenvalue problems. In *Applications of bifurcation theory (Proc. Advanced Sem., Univ. Wisconsin, Madison, Wis., 1976)*, pages 359–384. Publ. Math. Res. Center, No. 38. Academic Press, New York, 1977.

[77] C. S. Kenney and A. J. Laub. A Schur-Fréchet algorithm for computing the logarithm and exponential of a matrix. *SIAM J. Matrix Anal. Appl.*, 19(3):640–663, 1998.

[78] S.M. Kirkup and S. Amini. Solution of the helmholtz eigenvalue problem via the boundary element method. *International Journal for Numerical Methods in Engineering*, 36(2):321–330, 1993.

[79] M. Kitahara. *Boundary Integral Equation Methods in Eigenvalue Problems of Elastodynamics and Thin Plates*. Number 10 in Studies in applied mechanics. Elsevier, Amsterdam, 1985.

[80] A. V. Knyazev. Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method. *SIAM J. Sci. Comput.*, 23(2):517–541, 2001. Copper Mountain Conference (2000).

[81] T. Košir. Kronecker bases for linear matrix equations, with application to two-parameter eigenvalue problems. *Linear Algebra Appl.*, 249:259–288, 1996.

[82] A. Kostić and H. Voss. On Sylvester's law of inertia for nonlinear eigenvalue problems. *Electron. Trans. Numer. Anal.*, 40:82–93, 2013.

[83] D. Kressner. A block Newton method for nonlinear eigenvalue problems. *Numer. Math.*, 114(2):355–372, 2009.

[84] V. N. Kublanovskaya. On an approach to the solution of the generalized latent value problem for λ-matrices. *SIAM J. Numer. Anal.*, 7:532–537, 1970.

[85] P. Kuchment. *Floquet theory for partial differential equations*, volume 60 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1993.

[86] R. B. Lehoucq and D. C. Sorensen. Deflation techniques for an implicitly restarted Arnoldi iteration. *SIAM J. Matrix Anal. Appl.*, 17(4):789–821, 1996.

[87] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, volume 6 of *Software, Environments, and Tools*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.

[88] B.-S. Liao, Z. Bai, L.-Q. Lee, and K. Ko. Nonlinear Rayleigh-Ritz iterative method for solving large scale nonlinear eigenvalue problems. *Taiwanese J. Math.*, 14(3A):869–883, 2010.

[89] P. Lindqvist. On nonlinear Rayleigh quotients. *Potential Anal.*, 2(3):199–218, 1993.

[90] G. G. Lorentz. *Approximation of Functions*. Chelsea Publishing Co., New York, second edition, 1986.

[91] S. H. Lui, H. B. Keller, and T. W. C. Kwok. Homotopy method for the large, sparse, real nonsymmetric eigenvalue problem. *SIAM J. Matrix Anal. Appl.*, 18(2):312–333, 1997.

[92] M. Luo, Q. H. Liu, and Z. Li. Spectral element method for band structures of two-dimensional anisotropic photonic crystals. *Phys. Rev. E*, 79:026705, 2009.

[93] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Vector spaces of linearizations for matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 28(4):971–1004, 2006.

[94] R. Mathias. A chain rule for matrix functions and applications. *SIAM J. Matrix Anal. Appl.*, 17(3):610–620, 1996.

[95] K. Meerbergen. Locking and restarting quadratic eigenvalue solvers. *SIAM J. Sci. Comput.*, 22(5):1814–1839, 2000.

[96] V. Mehrmann and C. Schröder. Nonlinear eigenvalue and frequency response problems in industrial practice. *J. Math. Ind.*, 1(7):1–18, 2011.

[97] V. Mehrmann and H. Voss. Nonlinear eigenvalue problems: a challenge for modern eigenvalue methods. *GAMM Mitt. Ges. Angew. Math. Mech.*, 27(2):121–152, 2004.

[98] R. Mennicken and M. Möller. *Non-Self-Adjoint Boundary Eigenvalue Problems*, volume 192 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 2003.

[99] W. Michiels and S.-I. Niculescu. *Stability and Stabilization of Time-Delay Systems*, volume 12 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007. An eigenvalue-based approach.

[100] E. Moreno, D. Erni, and C. Hafner. Band structure computations of metallic photonic crystals with the multiple multipole method. *Phys. Rev. B*, 65(15):155120, 2002.

[101] A. Neumaier. Residual inverse iteration for the nonlinear eigenvalue problem. *SIAM J. Numer. Anal.*, 22(5):914–923, 1985.

[102] K. Neymeyr. A geometric theory for preconditioned inverse iteration. I. Extrema of the Rayleigh quotient. *Linear Algebra Appl.*, 322(1-3):61–85, 2001.

[103] K. Neymeyr. A geometric theory for preconditioned inverse iteration. II. Convergence estimates. *Linear Algebra Appl.*, 322(1-3):87–104, 2001.

[104] A.J. Nowak and C.A. Brebbia. The multiple-reciprocity method. A new approach for transforming BEM domain integrals to the boundary. *Engineering Analysis with Boundary Elements*, 6(3):164–167, 1989.

[105] F. Odeh and J. B. Keller. Partial differential equations with periodic coefficients and Bloch waves in crystals. *J. Mathematical Phys.*, 5:1499–1504, 1964.

[106] M. R. Osborne. A new method for the solution of eigenvalue problems. *Comput. J.*, 7:228–232, 1964.

[107] C. C. Paige, B. N. Parlett, and H. A. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numer. Linear Algebra Appl.*, 2(2):115–133, 1995.

[108] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. IV. Analysis of Operators*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1978.

[109] W. C. Rheinboldt. A unified convergence theory for a class of iterative processes. *SIAM J. Numer. Anal.*, 5:42–63, 1968.

[110] S. Rjasanow and O. Steinbach. *The Fast Solution of Boundary Integral Equations*. Mathematical and Analytical Techniques with Applications to Engineering. Springer, New York, 2007.

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

[112] A. Ruhe. Algorithms for the nonlinear eigenvalue problem. *SIAM J. Numer. Anal.*, 10:674–689, 1973.

[113] K. Sakoda, N. Kawai, T. Ito, A. Chutinan, S. Noda, T. Mitsuyu, and K. Hirao. Photonic bands of metallic systems. i. principle of calculation and accuracy. *Phys. Rev. B*, 64:045116, 2001.

[114] J. Sanchez Hubert and E. Sanchez-Palencia. *Vibration and Coupling of Continuous Systems*. Springer-Verlag, Berlin, 1989. Asymptotic methods.

[115] S. A. Sauter and C. Schwab. *Boundary Element Methods*, volume 39 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2011. Translated and expanded from the 2004 German original.

[116] K. Schmidt and P. Kauf. Computation of the band structure of two-dimensional photonic crystals with hp finite elements. *Computer Methods in Applied Mechanics and Engineering*, 198(13–14):1249–1259, 2009.

[117] H. Schwetlick and K. Schreiber. Nonlinear Rayleigh functionals. *Linear Algebra Appl.*, 436(10):3991–4016, 2012.

[118] K.V. Singh and Y.M. Ram. Transcendental eigenvalue problem and its applications. *AIAA Journal*, 40(7):1402–1407, 2002.

[119] G. L. G. Sleijpen and H. A. Van der Vorst. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM Rev.*, 42(2):267–293, 2000.

[120] G. L. G. Sleijpen, H. A. van der Vorst, and E. Meijerink. Efficient expansion of subspaces in the Jacobi-Davidson method for standard and generalized eigenproblems. *Electron. Trans. Numer. Anal.*, 7:75–89, 1998. Large scale eigenvalue problems (Argonne, IL, 1997).

[121] S. I. Solov'ëv. Eigenvibrations of a plate with elastically attached load. Preprint SFB393/03-06, Chemnitz University of Technology, 2003.

[122] S. I. Solov'ëv. Preconditioned iterative methods for a class of nonlinear eigenvalue problems. *Linear Algebra Appl.*, 415(1):210–229, 2006.

[123] O. Steinbach. *Numerical Approximation Methods for Elliptic Boundary Value Problems*. Springer, New York, 2008. Finite and boundary elements, Translated from the 2003 German original.

[124] O. Steinbach and G. Unger. A boundary element method for the Dirichlet eigenvalue problem of the Laplace operator. *Numer. Math.*, 113(2):281–298, 2009.

[125] G. W. Stewart. A Krylov-Schur algorithm for large eigenproblems. *SIAM J. Matrix Anal. Appl.*, 23(3):601–614, 2001/02.

[126] B. Sz.-Nagy. On a spectral problem of Atkinson. *Acta Math. Acad. Sci. Hungar.*, 3:61–66, 1952.

[127] F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Rev.*, 43(2):235–286, 2001.

[128] G. Unger. Private communication, 2011.

[129] G. Unger. Convergence orders of iterative methods for nonlinear eigenvalue problems. In *Advanced Finite Element Methods and Applications*, volume 66 of *Lect. Notes Appl. Comput. Mech.*, pages 217–237. Springer, Heidelberg, 2013.

[130] R. Van Beeumen, K. Meerbergen, and W. Michiels. A rational Krylov method based on Hermite interpolation for nonlinear eigenvalue problems. *SIAM J. Sci. Comput.*, 35(1):A327–A350, 2013.

[131] H. Voss. Initializing iterative projection methods for rational symmetric eigenproblems. In *Online Proceedings of the Dagstuhl Seminar on Theoretical and Computational Aspects of Matrix Algorithms*, Schloss Dagstuhl, 2003.

[132] H. Voss. A maxmin principle for nonlinear eigenvalue problems with application to a rational spectral problem in fluid-solid vibration. *Appl. Math.*, 48(6):607–622, 2003. Mathematical and computer modeling in science and engineering.

[133] H. Voss. A rational spectral problem in fluid-solid vibration. *Electron. Trans. Numer. Anal.*, 16:93–105, 2003.

[134] H. Voss. An Arnoldi method for nonlinear eigenvalue problems. *BIT*, 44(2):387–401, 2004.

[135] H. Voss. Eigenvibrations of a plate with elastically attached loads. In P. Neittaanmäki, T. Rossi, S. Korotov, E. Oñate, J. Périaux, and D. Knörzer, editors, *Proceedings of the European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS)*, Jyväskylä, 2004.

[136] H. Voss. Iterative projection methods for computing relevant energy states of a quantum dot. *J. Comput. Phys.*, 217(2):824–833, 2006.

[137] H. Voss. A Jacobi-Davidson method for nonlinear and nonsymmetric eigenproblems. *Comput. & Structures*, 85(17-18):1284–1292, 2007.

[138] H. Voss. A minmax principle for nonlinear eigenproblems depending continuously on the eigenparameter. *Numer. Linear Algebra Appl.*, 16(11-12):899–913, 2009.

[139] H. Voss and B. Werner. A minimax principle for nonlinear eigenvalue problems with applications to nonoverdamped systems. *Math. Methods Appl. Sci.*, 4(3):415–424, 1982.

[140] H. Voss and B. Werner. Solving sparse nonlinear eigenvalue problems. Technical Report 82/4, Institut für Angewandte Mathematik, Universität Hamburg, 1982.

[141] J. Wilkening. An algorithm for computing Jordan chains and inverting analytic matrix functions. *Linear Algebra Appl.*, 427(1):6–25, 2007.

[142] J. Wu. *Theory and Applications of Partial Functional-Differential Equations*, volume 119 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.

[143] Y. P. Zhigalko, A. D. Lyashko, and S. I. Solov'ëv. Vibrations of a cylindrical shell with joined rigid annular elements. *Model. Mekh.*, 2:68–85, 1988.

[144] Y. P. Zhigalko and S. I. Solov'ëv. Natural oscillations of a beam with a harmonic oscillator. *Russian Math.*, 45(10):33–35, 2001.

Curriculum Vitae

Personal information

Name:	Effenberger, Cedric
Date of birth:	06.09.1983
Place of birth:	Stollberg, Germany
Nationality:	German

Education

since 03/2012	**École polytechnique fédérale de Lausanne** *Doctoral studies in Computational Science and Engineering* Research topic: Robust solution methods for nonlinear eigenvalue problems
11/2009 – 02/2012	**Eidgenössische Technische Hochschule Zürich** *Doctoral studies in Computational Science and Engineering* Research topic: Robust solution methods for nonlinear eigenvalue problems
08/2009 – 10/2009	**Technische Universität Chemnitz** *Research assistant*
10/2003 – 07/2009	**Technische Universität Chemnitz** *Diploma studies in Industrial Mathematics* Topic of diploma thesis: Rational Krylov subspace methods for Hamiltonian eigenproblems
07/2002	**Johannes-Kepler-Gymnasium Chemnitz** *German Abitur*

Awards

10/2009	Award of the German Mathematical Society for outstanding diploma thesis
06/2002	Award of the Saxonian ministry of education for outstanding high school graduation

Conferences and seminars

03/2013 **KU Leuven, Leuven, Belgium**
invited talk on *Robust successive computation of eigenpairs for nonlinear eigenvalue problems*

12/2012 **TU Hamburg-Harburg, Hamburg, Germany**
invited talk on *Robust successive computation of eigenpairs for nonlinear eigenvalue problems*

06/2012 **SIAM Conference on Applied Linear Algebra in Valencia, Spain**
invited minisymposium talk on *Robust successive computation of eigenpairs for nonlinear eigenvalue problems*

08/2011 **ILAS Conference in Braunschweig, Germany**
invited minisymposium talk on *Polynomial approximations to a class of nonlinear (PDE) eigenvalue problems*

04/2011 **GAMM Annual Meeting in Graz, Austria**
invited minisymposium talk on *Continuation of Eigenvalues and Invariant Pairs for Parameterized Nonlinear Eigenvalue Problems*

10/2009 **Studierendenkonferenz der Deutschen Mathematiker-Vereinigung, Bochum, Germany**
contributed talk on *Rational Krylov Subspace Methods for Hamiltonian Eigenproblems*

09/2008 **GAMM Workshop on Applied and Numerical Linear Algebra, Hamburg-Harburg, Germany**
contributed talk on *Rational Krylov Methods for the Hamiltonian Eigenvalue Problem*

Publications

[1] C. Effenberger. *Robust successive computation of eigenpairs for nonlinear eigenvalue problems.* To appear in SIAM Journal on Matrix Analysis and Applications.

[2] C. Effenberger, D. Kressner, O. Steinbach, and G. Unger. *Interpolation-based solution of a nonlinear eigenvalue problem in fluid-structure interaction.* Proceedings in Applied Mathematics and Mechanics, 12:633–634, 2012.

[3] C. Effenberger and D. Kressner. *Chebyshev interpolation for nonlinear eigenvalue problems.* BIT Numerical Mathematics, 52(4):933–951, 2012.

[4] C. Effenberger, D. Kressner, and C. Engström. *Linearization techniques for band structure calculations in absorbing photonic crystals.* International Journal for Numerical Methods in Engineering, 89(2):180–191, 2012.

[5] W.-J. Beyn, C. Effenberger, and D. Kressner. *Continuation of eigenvalues and invariant pairs for parameterized nonlinear eigenvalue problems.* Numerische Mathematik, 119:489–516, 2011.

[6] C. Effenberger. *Rational Krylov subspace methods for Hamiltonian eigenproblems.* Diploma thesis, Chemnitz University of Technology, 2009. supervised by Prof. Dr. Peter Benner.

[7] P. Benner and C. Effenberger. *A rational SHIRA method for the Hamiltonian eigenvalue problem.* Taiwanese Journal of Mathematics, 14(3A):805–823, 2010.