115
Nonlinear Eigenvalue Problems

Heinrich Voss
Hamburg University of Technology

This chapter considers the nonlinear eigenvalue problem to find a parameter λ such that the linear system

$$T(\lambda)\mathbf{x} = 0 \tag{115.1}$$

has a nontrivial solution $\mathbf{x}$, where $T(\cdot) : D \to \mathbb{C}^{n\times n}$ is a family of matrices depending on a complex parameter $\lambda \in D$.

It generalizes the linear eigenvalue problem $A\mathbf{x} = \lambda\mathbf{x}$, $A \in \mathbb{C}^{n\times n}$, where $T(\lambda) = \lambda I - A$, and the generalized linear eigenvalue problem where $T(\lambda) = \lambda B - A$, $A, B \in \mathbb{C}^{n\times n}$.

Nonlinear eigenvalue problems $T(\lambda)\mathbf{x} = 0$ arise in a variety of applications in science and engineering, such as the dynamic analysis of structures, vibrations of fluid–solid structures, the electronic behavior of quantum dots, and delay eigenvalue problems, to name just a few. Due to its wide range of applications, the quadratic eigenvalue problem $T(\lambda)\mathbf{x} = \lambda^2 M\mathbf{x} + \lambda C\mathbf{x} + K\mathbf{x} = 0$ is of particular interest, but also polynomial, rational and more general eigenvalue problems appear. A standard approach for investigating or numerically solving polynomial eigenvalue problems is linearization where the original problem is transformed into a generalized linear eigenvalue problem with the same spectrum. Details on linearization and structure preservation are discussed in Chapter 102, Matrix Polynomials.

This chapter is concerned with the general nonlinear eigenvalue problem which in general can not be linearized. Unlike for linear and polynomial eigenvalue problems there may exist infinitely many eigenvalues. In practice, however, one is usually interested only in a few eigenvalues close to a target value or a line in the complex plane.

If T is linear then $T(\lambda) = T(0) + \lambda T'(0)$ has the form of a generalized eigenvalue problem, and in the general case linerization gives the approximation $T(\lambda) = T(0) + \lambda T'(0) + O(\lambda^2)$, which is again a generalized linear eigenvalue problem. Hence, it is not surprising, that the (elementwise) derivative $T'(\lambda)$ of $T(\lambda)$ plays an important role in the analysis of nonlinear eigenvalue problems.

We tacitly assume in the whole chapter that whenever a derivative $T'(\hat{\lambda})$ appears, T is analytic in a neighborhood of $\hat{\lambda}$ or in the real case $T : D \to \mathbb{R}^{n\times n}$, $D \subset \mathbb{R}$ that T is differentiable in a neighborhood of $\hat{\lambda}$. $\|\cdot\|$ always denotes the Euclidean and spectral norm, respectively, and we use the notation $[\mathbf{x};\mathbf{y}] := [\mathbf{x}^T, \mathbf{y}^T]^T$ for column vectors.

115.1 Basic Properties

This section presents basic properties of the nonlinear eigenvalue problem (115.1)

Definitions:

As for a linear eigenvalue problem, $\hat{\lambda} \in D$ is called an **eigenvalue** of $T(\cdot)$ if $T(\hat{\lambda})\mathbf{x} = 0$ has a nontrivial solution $\hat{\mathbf{x}} \neq 0$. Then $\hat{\mathbf{x}}$ is called a corresponding **eigenvector** or **right eigenvector**, and $(\hat{\lambda}, \hat{\mathbf{x}})$ is called **eigenpair** of $T(\cdot)$.

Any nontrivial solution $\hat{\mathbf{y}} \neq 0$ of the adjoint equation $T(\hat{\lambda})^*\mathbf{y} = 0$ is called **left eigenvector** of $T(\cdot)$ and the vector-scalar-vector triplet $(\hat{\mathbf{y}}, \hat{\lambda}, \hat{\mathbf{x}})$ is called **eigentriplet** of $T(\cdot)$.

The eigenvalue problem (115.1) is **regular** if $\det T(\lambda) \not\equiv 0$, and otherwise it is called **singular**.

The **spectrum** $\sigma(T(\cdot))$ of $T(\cdot)$ is the set of all eigenvalues of $T(\cdot)$.

An eigenvalue $\hat{\lambda}$ of $T(\cdot)$ has **algebraic multiplicity** k if $\frac{d^\ell}{d\lambda^\ell} \det(T(\lambda))\Big|_{\lambda=\hat{\lambda}} = 0$ for $\ell = 0, \ldots, k-1$ and $\frac{d^k}{d\lambda^k} \det(T(\lambda))\Big|_{\lambda=\hat{\lambda}} \neq 0$.

An eigenvalue $\hat{\lambda}$ is **simple** if its algebraic multiplicity is one.

The **geometric multiplicity** of an eigenvalue $\hat{\lambda}$ is the dimension of the kernel $\ker(T(\hat{\lambda}))$ of $T(\hat{\lambda})$.

An eigenvalue $\hat{\lambda}$ is called **semi-simple** if its algebraic and geometric mutiplicity coincide.

$T(\cdot) : J \to \mathbb{R}^{n\times n}$ is **real symmetric** if $T(\lambda) = T(\lambda)^T$ for every $\lambda \in J \subset \mathbb{R}$.

$T(\cdot) : D \to \mathbb{C}^{n\times n}$ is **complex symmetric** if $T(\lambda) = T(\lambda)^T$ for every $\lambda \in D$.

$T(\cdot) : D \to \mathbb{C}^{n\times n}$ is **Hermitian** if D is symmetric with respect to the real line and $T(\lambda)^* = T(\bar{\lambda})$ for every $\lambda \in D$.

Facts:

1. For $A \in \mathbb{C}^{n\times n}$ and $T(\lambda) = \lambda I - A$, the terms eigenvalue, (left and right) eigenvector, eigenpair, eigentriplet, spectrum, algebraic and geometric multiplicity and semi-simple have their standard meaning.
2. For linear eigenvalue problems,
   - eigenvectors corresponding to distinct eigenvalues are linearly independent, which is not the case for nonlinear eigenvalue problems (cf. Example 1).
   - left and right eigenvectors corresponding to distinct eigenvalues are orthogonal, which does not hold for nonlinear eigenproblems (cf. Example 2).
   - the algebraic multiplicities of eigenvalues sum up to the dimension of the problem, whereas for nonlinear problems there may exist an infinite number of eigenvalues (cf. Example 2) and an eigenvalue may have any algebraic multiplicity (cf. Example 3).
3. [Sch08] If $\hat{\lambda}$ is an algebraically simple eigenvalue of $T(\cdot)$, then $\hat{\lambda}$ is geometrically simple.
4. [Neu85, Sch08] Let $(\hat{\mathbf{y}}, \hat{\lambda}, \hat{\mathbf{x}})$ be an eigentriplet of $T(\cdot)$. Then $\hat{\lambda}$ is algebraically simple if and only if $\hat{\lambda}$ is geometrically simple and $\hat{\mathbf{y}}^* T'(\hat{\lambda})\hat{\mathbf{x}} \neq 0$.
5. [Sch08] Let $D \subset \mathbb{C}$ and $E \subset \mathbb{C}^d$ be open sets. Let $T : D \times E \to \mathbb{C}^{n\times n}$ be continuously differentiable, and let $\hat{\lambda}$ be a simple eigenvalue of $T(\cdot, 0)$ and $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ right and left eigenvectors with unit norm. Then the first order perturbation expansion at $\hat{\lambda}$ reads as follows:

$$\lambda(\varepsilon) - \hat{\lambda} = \frac{1}{\hat{\mathbf{y}}^* T'(\hat{\lambda}, 0)\hat{\mathbf{x}}} \sum_{j=1}^{d} \varepsilon_j \hat{\mathbf{y}}^* \frac{\partial T}{\partial \varepsilon_j}(\hat{\lambda}, 0)\hat{\mathbf{x}} + o(\|\varepsilon\|).$$

The normwise condition number for $\hat{\lambda}$ is given by

$$\kappa(\hat{\lambda}) = \limsup_{\|\varepsilon\|\to 0} \frac{|\lambda(\varepsilon) - \hat{\lambda}|}{\|\varepsilon\|} = \frac{1}{|\hat{\mathbf{y}}^* T'(\hat{\lambda}, 0)\hat{\mathbf{x}}|} \sqrt{\sum_{j=1}^{d} \left| \hat{\mathbf{y}}^* \frac{\partial T}{\partial \varepsilon_j}(\hat{\lambda}, 0)\hat{\mathbf{x}} \right|^2}.$$

6. [Sch08] Let $(\hat{\mathbf{y}}, \hat{\lambda}, \hat{\mathbf{x}})$ be an eigentriplet of $T(\cdot)$ with simple eigenvalue $\hat{\lambda}$. Then for sufficiently small $|\hat{\lambda} - \lambda|$

$$T(\lambda)^{-1} = \frac{1}{\lambda - \hat{\lambda}} \frac{\hat{\mathbf{x}}\hat{\mathbf{y}}^*}{\hat{\mathbf{y}}^* T'(\hat{\lambda})\hat{\mathbf{x}}} + O(1).$$

7. [Neu85] Let $\hat{\lambda}$ be a simple eigenvalue of $T(\cdot)$, and let $\hat{\mathbf{x}}$ be a right eigenvector normalized such that $\mathbf{e}^*\hat{\mathbf{x}} = 1$ for some vector $\mathbf{e}$. Then the matrix $B := T(\hat{\lambda}) + T'(\hat{\lambda})\hat{\mathbf{x}}\mathbf{e}^*$ is nonsingular.
8. If $T(\cdot)$ is real symmetric and λ is a real eigenvalue, then left and right eigenvectors corresponding to λ coincide.
9. If $T(\cdot)$ is complex symmetric and $\mathbf{x}$ is a right eigenvector, then $\bar{\mathbf{x}}$ is a left eigenvector corresponding to the same eigenvalue.
10. If $T(\cdot)$ is Hermitian, then eigenvalues are real (and left and right eigenvectors corresponding to λ coincide) or they come in pairs, i.e. if $(\mathbf{y}, \lambda, \mathbf{x})$ is an eigentriplet of $T(\cdot)$, then this is also true for $(\mathbf{x}, \bar{\lambda}, \mathbf{y})$.

Examples:

1. For the quadratic eigenvalue problem $T(\lambda)\mathbf{x} = 0$ with

$$T(\lambda) := \begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix} + \lambda \begin{bmatrix} 7 & -5 \\ 10 & -8 \end{bmatrix} + \lambda^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{115.2}$$

the distinct eigenvalues $\lambda = 1$ and $\lambda = 2$ share the eigenvector $[1; 2]$.

2. Let $T(\lambda)\mathbf{x} := \begin{bmatrix} e^{i\lambda^2} & 1 \\ 1 & 1 \end{bmatrix} \mathbf{x} = 0$. Then $T(\lambda)\mathbf{x} = 0$ has a countable set of eigenvalues $\sqrt{2k\pi}$, $k \in \mathbb{N} \cup \{0\}$. $\hat{\lambda} = 0$ is an algebraically double and geometrically simple eigenvalue with left and right eigenvectors $\hat{\mathbf{x}} = \hat{\mathbf{y}} = [1; -1]$, and $\hat{\mathbf{y}}^* T'(0)\hat{\mathbf{x}} = 0$. Every $\hat{\lambda}_k = \sqrt{2k\pi}$, $k \neq 0$ is algebraically and geometrically simple with the same eigenvectors $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ as before, and $\hat{\mathbf{y}}^* T'(\hat{\lambda}_k)\hat{\mathbf{x}} = 2\sqrt{2k\pi}i \neq 0$.
3. $T(\lambda) = (\lambda^k)$, $k \in \mathbb{N}$ has the eigenvalue $\hat{\lambda} = 0$ with algebraic multiplicity k.

115.2 Analytic matrix functions

In this section we consider the eigenvalue problem (115.1) where $T(\cdot) : D \to \mathbb{C}^{n\times n}$ is a regular matrix function which is analytic in a neighborhood of an eigenvalue $\hat{\lambda}$.

Definitions:

A sequence of vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{r-1}$ is called a **Jordan chain** (of length r) corresponding to $\hat{\lambda}$ if $\mathbf{x}_0 \neq 0$ and

$$\sum_{k=0}^{\ell} \frac{1}{k!} \frac{d^k T(\hat{\lambda})}{d\lambda^k}\bigg|_{\lambda=\hat{\lambda}} \mathbf{x}_{\ell-k} = 0 \quad \text{for } \ell = 0, \ldots, r-1.$$

$\mathbf{x}_0$ is an **eigenvector** and $\mathbf{x}_1, \ldots, \mathbf{x}_{r-1}$ are **generalized eigenvectors**.

Let $\mathbf{x}_0$ be an eigenvector corresponding to an eigenvalue $\hat{\lambda}$. The maximal length of a Jordan chain that starts with $\mathbf{x}_0$ is called the **multiplicity of** $\mathbf{x}_0$.

An eigenvalue $\hat{\lambda}$ is is said to be **normal** if it is a discrete point in $\sigma(T(\cdot)))$ and the multiplicity of each corresponding eigenvector is finite.

An analytic function $\mathbf{x} : D \to \mathbb{C}^n$ is called **root function** of $T(\cdot)$ at $\hat{\lambda} \in D$ if $T(\hat{\lambda})\mathbf{x}(\hat{\lambda}) = 0$ and $\mathbf{x}(\hat{\lambda}) \neq 0$.

The multiplicity of $\hat{\lambda}$ as a zero of $T(\lambda)\mathbf{x}(\lambda)$ is called the **multiplicity** of $\mathbf{x}(\cdot)$.

The **rank** of an eigenvector $\mathbf{x}_0$ is the maximum of the multiplicities of all root functions $\mathbf{x}(\cdot)$ such that $\mathbf{x}(\hat{\lambda}) = \mathbf{x}_0$.

A root function $\mathbf{x}(\cdot)$ is called a **maximal root function** if the multiplicity of $\mathbf{x}(\cdot)$ is equal to the rank of $\mathbf{x}_0 := \mathbf{x}(\hat{\lambda})$.

Let $\mathbf{x}_0^{(1)} \in \ker T(\hat{\lambda})$ be an eigenvector with maximal rank and let $\mathbf{x}^{(1)}(\lambda) = \sum_{j=0}^{\infty} \mathbf{x}_j^{(1)}(\lambda - \hat{\lambda})^j$ be a maximal root function such that $\mathbf{x}^{(1)}(\hat{\lambda}) = \mathbf{x}_0^{(1)}$. Suppose that the root functions $\mathbf{x}^{(k)}(\lambda) = \sum_{j=0}^{\infty} \mathbf{x}_j^{(k)}(\lambda - \hat{\lambda})^j$, $k = 1, \ldots, i-1$ are already constructed, and let $\mathbf{x}_0^{(i)}$ be an eigenvector with maximal rank in some direct complement to the linear span of the vectors $\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_0^{(i-1)}$ in $\ker T(\hat{\lambda})$. Let $\mathbf{x}^{(i)}(\lambda) = \sum_{j=0}^{\infty} \mathbf{x}_j^{(i)}(\lambda - \hat{\lambda})^j$ be a maximal root function such that $\mathbf{x}^{(i)}(\hat{\lambda}) = \mathbf{x}_0^{(i)}$. Then the ordered set

$$\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_{r_1-1}^{(1)}, \mathbf{x}_0^{(2)}, \ldots, \mathbf{x}_{r_2-1}^{(2)}, \ldots, \mathbf{x}_0^{(k)}, \ldots, \mathbf{x}_{r_k-1}^{(k)},$$

where $k = \dim \ker T(\hat{\lambda})$ and $r_j = \operatorname{rank} \mathbf{x}_0^{(j)}$ is called **canonical set of Jordan chains**, and the ordered set $\mathbf{x}^{(1)}(\lambda), \ldots, \mathbf{x}^{(k)}(\lambda)$ is called **canonical system of root functions**.

Let $X \in \mathbb{C}^{n \times \alpha}$ contain in its columns the vectors of a canonical set of Jordan chains and let $J = \operatorname{diag}(J_1, \ldots, J_k)$, where J_j is a Jordan block of size $r_j \times r_j$ corresponding to $\hat{\lambda}$. Then the pair (X, J) is called a **Jordan pair**.

Let $\mathbf{x}^{(1)}(\lambda), \ldots, \mathbf{x}^{(k)}(\lambda)$ be a canonical system of root functions at $\hat{\lambda}$, and let $\mathbf{x}^{(k+1)}, \ldots, \mathbf{x}^{(n)} \in \mathbb{C}^n$ such that $\mathbf{x}^{(1)}(\hat{\lambda}), \ldots, \mathbf{x}^{(k)}(\hat{\lambda}), \mathbf{x}^{(k+1)}, \ldots, \mathbf{x}^{(n)}$ is a basis of $\mathbb{C}^n$. Then the system $\mathbf{x}^{(1)}(\lambda)$, $\ldots, \mathbf{x}^{(k)}(\lambda), \mathbf{x}^{(k+1)}, \ldots, \mathbf{x}^{(n)}$ is called an **extended canonical system of root functions**. To the constant functions $\mathbf{x}^{(k+1)}, \ldots, \mathbf{x}^{(n)} \in \mathbb{C}^n$ (which are not root functions in the strict sense of the definition) is assigned the multiplicity 0.

Let $\hat{\lambda}$ be an eigenvalue of $T(\cdot)$, and let $\Phi(\cdot)$ be an analytic matrix function such that its columns form an extended canonical system of root functions of $T(\cdot)$ at $\hat{\lambda}$. Then (cf. [GKS93]) in a neighborhood of $\hat{\lambda}$,

$$L(\lambda)\Phi(\lambda) = P(\lambda)D(\lambda), \tag{115.3}$$

where $D(\lambda)$ is a diagonal matrix with diagonal entries $(\lambda - \hat{\lambda})^{\kappa_1}, \ldots, (\lambda - \hat{\lambda})^{\kappa_n}$ and $P(\cdot)$ is a matrix function analytic at $\hat{\lambda}$ such that $\det P(\hat{\lambda}) \neq 0$. Furthermore, the exponents $\kappa_1, \ldots, \kappa_n$ are the multiplicities of the columns of $\Phi(\cdot)$, also called **partial multiplicities** of $T(\cdot)$ at $\hat{\lambda}$. (115.3) is the **local Smith form** of $T(\cdot)$ in a neighborhood of $\hat{\lambda}$.

A pair of matrices $(Y, Z) \in \mathbb{C}^{n \times p} \times \mathbb{C}^{p \times p}$ is a **regular pair** if for some integer $\ell \geq 1$

$$\operatorname{rank} \begin{bmatrix} Y \\ YZ \\ \vdots \\ YZ^{\ell-1} \end{bmatrix} = p.$$

The number p is called the **order** of the regular pair (Y, Z).

Facts:

The following facts for which no specific reference is given can be found in [GLR82, GR81].

1. In contrast to linear eigenvalue problems the vectors in a Jordan chain need not be linearly independent. Even the zero vector is admissible as a generalized eigenvector.
2. Let $\mathbf{x}(\cdot)$ be a root function at $\hat{\lambda}$, and let $\mathbf{x}^{(j)}$ denote the jth derivative of $\mathbf{x}$. Then the vectors $\mathbf{x}_j := \mathbf{x}^{(j)}(\hat{\lambda})$, $j = 0, \ldots, r-1$ form a Jordan chain at $\hat{\lambda}$, where r denotes the multiplicity of $\mathbf{x}(\cdot)$.
3. The multiplicity of a root function at $\hat{\lambda}$ (and hence the rank of an eigenvector) is at most the algebraic multiplicity of $\hat{\lambda}$.
4. The numbers $r_1 \geq \cdots \geq r_k$ in a Jordan pair are uniquely determined.
5. The number $\alpha := r_1 + \cdots + r_k$ is the algebraic multiplicity of the eigenvalue $\hat{\lambda}$.
6. [GKS93] Let $\mathbf{y}_1, \ldots, \mathbf{y}_\ell : D \to \mathbb{C}^n$ be a set of root functions at $\hat{\lambda}$ with multiplicities $s_1 \geq \cdots \geq s_\ell$ such that $\mathbf{y}_1(\hat{\lambda}), \ldots, \mathbf{y}_\ell(\hat{\lambda}) \in \ker T(\hat{\lambda})$ are linearly independent. If the root functions $\mathbf{x}_1, \ldots, \mathbf{x}_k$ define a canonical set of Jordan chains of $T(\cdot)$ at $\hat{\lambda}$ with multiplicities $r_1 \geq \cdots \geq r_k$, then $k \geq \ell$ and $r_i \geq s_i$ for $i = 1, \ldots, \ell$. Moreover, $\mathbf{y}_1, \ldots, \mathbf{y}_\ell$ define a canonical set of Jordan chains of $T(\cdot)$ at $\hat{\lambda}$ if and only if $\ell = k$ and $s_j = r_j$ for $j = 1, \ldots, \ell$.
7. Let $S(\cdot)$ be an analytic matrix function with $\det S(\hat{\lambda}) \neq 0$. $\mathbf{x}_0, \ldots, \mathbf{x}_k$ is a Jordan chain of $T(\cdot)S(\cdot)$ corresponding to $\hat{\lambda}$ if and only if the vectors $\mathbf{y}_0, \ldots, \mathbf{y}_k$ given by $\mathbf{y}_j = \sum_{i=0}^{j} \frac{1}{i!} S^{(i)}(\hat{\lambda}) \mathbf{x}_{j-i}$, $j = 0, \ldots, k-1$ is a Jordan chain of $T(\cdot)$ corresponding to $\hat{\lambda}$.
8. For S as in the last fact the Jordan chains of $T(\cdot)$ coincide with those of $S(\cdot)T(\cdot)$ corresponding to the same $\hat{\lambda}$.
9. Two regular analytic matrix functions $T_1(\cdot)$ and $T_2(\cdot)$ have a common Jordan pair at $\hat{\lambda}$ if and only if $T_2(\lambda)T_1^{-1}(\lambda)$ is analytic and invertible at $\hat{\lambda}$.
10. [GKS93] Let $T(\cdot)$, $\Phi(\cdot)$, $D(\cdot)$ and $P(\cdot)$ be regular $n \times n$ matrix functions, analytic at $\hat{\lambda}$, such that $L(\lambda)\Phi(\lambda) = P(\lambda)D(\lambda)$ in a neighborhood of $\hat{\lambda}$. Assume that $\Phi(\hat{\lambda})$ is invertible and that $D(\cdot)$ is a diagonal matrix polynomial with diagonal entries $(\lambda - \hat{\lambda})^{\kappa_1}, \ldots, (\lambda - \hat{\lambda})^{\kappa_n}$, where $\kappa_1 \geq \cdots \geq \kappa_n$. Then the following three conditions are equivalent:

    (i) the columns of $\Phi(\cdot)$ form an extended canonical system of root functions of $T(\cdot)$ at $\hat{\lambda}$ with partial multiplicities $\kappa_1, \ldots, \kappa_n$

    (ii) $\det P(\hat{\lambda}) \neq 0$

    (iii) $\sum_{j=1}^{n} \kappa_j$ is the the algebraic multiplicity of $\hat{\lambda}$.

11. [GKS93] Let $\mathbf{x}(\lambda) = \sum_{j=0}^{\infty} (\lambda - \hat{\lambda})^j \mathbf{x}_j$ be an analytic $\mathbb{C}^n$-vector function with $\mathbf{x}_0 \neq 0$, and set $X := [\mathbf{x}_0, \ldots, \mathbf{x}_p]$. Then $\mathbf{x}(\cdot)$ is a root function of $T(\cdot)$ at $\hat{\lambda}$ of multiplicity at most p if and only if $T(\lambda)X(\lambda I - J_{\hat{\lambda},p})^{-1}$ is an $n \times p$ analytic matrix function. Here $J_{\hat{\lambda},p}$ denotes a $p \times p$ Jordan block with eigenvalue $\hat{\lambda}$.
12. [AST09] $T(\cdot)$ admits a representation $P(\lambda)T(\lambda)Q(\lambda) = D(\lambda)$ where $P(\cdot)$ and $Q(\cdot)$ are regular analytic matrix functions with constant nonzero determinants, and $D(\lambda) = \operatorname{diag}\{d_j(\lambda)\}_{j=1\ldots,n}$ is a diagonal matrix of analytic functions such that $d_j(\lambda)/d_{j-1}(\lambda)$ are analytic for $j = 2, 3, \ldots, n$. This representation is also called local Smith form.
13. [AST09] With the representation in the last fact, if $\mathbf{q}_j(\lambda)$ is the jth column of Q, and $\hat{\lambda}$ a zero of $d_j(\cdot)$, then $(\hat{\lambda}, \mathbf{q}_j(\hat{\lambda}))$ is an eigenpair of $T(\cdot)$.
14. The non-zero partial multiplicities κ_j in the local Smith form of $T(\cdot)$ at $\hat{\lambda}$ coincide with the lengths $r_1 \geq \cdots \geq r_k$ of Jordan chains in a canonical set.
15. A Jordan pair (X, J) of $T(\cdot)$ at an eigenvalue $\hat{\lambda}$ is regular.
16. [GR81] Let $\hat{\lambda}$ be an eigenvalue of $T(\cdot)$ with algebraic multiplicity α, and let $(Y, Z) \in \mathbb{C}^{n\times\alpha} \times \mathbb{C}^{\alpha\times\alpha}$ be a pair of matrices such that $\sigma(Z) = \{\hat{\lambda}\}$. (Y, Z) is similar to a Jordan

pair (X, J) (i.e. $Y = XS$ and $Z = S^{-1}JS$ for some invertible matrix S) if and only if (Y, Z) is regular and the following equation holds:

$$\sum_{i=0}^{\infty} T_j Y (T - \hat{\lambda} I)^j = 0, \quad \text{where } T_j = \frac{1}{j!} T^{(j)}(\hat{\lambda})$$

(note that only a finite number of terms in the left-hand side of the equation is different from zero, because $\sigma(Z) = \{\hat{\lambda}\}$).

17. [HL99] Suppose that $A(\lambda)$ and $B(\lambda)$ are analytic matrix-valued functions such that $A(\hat{\lambda})$ and $B(\hat{\lambda})$ are non-singular. Then the partial multiplicities of the eigenvalue $\hat{\lambda}$ of $T(\lambda)$ and $\tilde{T}(\lambda) := B(\lambda)A(\lambda)C(\lambda)$ coincide.
18. [HL99] Suppose that a matrix-valued function $T(\lambda, \tau)$ depends analytically on λ and continuously on τ and that $\hat{\lambda} = 0$ is an eigenvalue of $T(\cdot, 0)$ of algebraic multiplicity α. Then there exists a neighborhood $\mathcal{O}$ of $\hat{\lambda}$ such that, for all τ sufficiently close to the origin, there are exactly α eigenvalues (counting with algebraic multiplicities) of the matrix-valued function $T(\cdot, \tau)$ in $\mathcal{O}$.

Examples:

1. [GLR82] For $T(\lambda) = \begin{bmatrix} \lambda^2 & -\lambda \\ 0 & \lambda^2 \end{bmatrix}$ we have $\det T(\lambda) = \lambda^4$, and hence $\hat{\lambda} = 0$ is an eigenvalue of $T(\cdot)$ with algebraic multiplicity 4 and geometric multiplicity 2.
   For an eigenvector $\mathbf{x}_0 = [x_{01}; x_{02}]$ the first generalized eigenvector $\mathbf{x}_1$ satisfies $T'(0)\mathbf{x}_0 + T(0)\mathbf{x}_1 = 0$, i.e. $\begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \mathbf{x}_0 = 0$, and $\mathbf{x}_1$ exists if and only if $x_{02} = 0$, and $\mathbf{x}_1$ can be taken completely arbitrary. For a second generalized eigenvector $\mathbf{x}_2$ we have $\frac{1}{2}T''(0)\mathbf{x}_0 + T'(0)\mathbf{x}_1 + T(0)\mathbf{x}_2 = 0$, i.e. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}_0 + \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \mathbf{x}_1 = 0$, i.e. $x_{01} = x_{12}$, and if this equation is satisfied, $\mathbf{x}_2$ can be chosen arbitrarily. The condition for the third generalized eigenvector $\mathbf{x}_3$ reads $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}_1 + \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \mathbf{x}_2 = 0$, which implies $x_{12} = 0$ and is contradictory.
   To summarize, the length of a Jordan chain can not exceed 3. Jordan chains of length 1 are $\mathbf{x}_0$, $\mathbf{x}_0 \neq 0$, Jordan chains of length 2 are $\mathbf{x}_0 = [x_{01}; 0], \mathbf{x}_1$ with $x_{01} \neq 0$ and $\mathbf{x}_1$ arbitrary, and Jordan chains of length 3 are $\mathbf{x}_0 = [x_{01}; 0], \mathbf{x}_1 = [x_{11}; x_{01}], \mathbf{x}_2$, where $x_{01} \neq 0$, and x_{11} and x_2 are arbitrary. One example of a canonical system of Jordan chains is $\mathbf{x}_0^{(1)} = [1; 0]$, $\mathbf{x}_1^{(1)} = [0; 1]$, $\mathbf{x}_2^{(1)} = [1; 1]$, $\mathbf{x}_0^{(2)} = [0; 1]$.
   $T(0) = 0$ implies that $\mathbf{x}(\cdot)$ is a root function at $\hat{\lambda} = 0$, if x_1 and x_2 are analytic and $\mathbf{x}(0) \neq 0$. $T(\lambda)\mathbf{x}(\lambda) = [\lambda^2 x_1(\lambda) - \lambda x_2(\lambda); \lambda^2 x_2(\lambda)] = 0$ yields that $\mathbf{x}$ has at least the multiplicity 2, and if $x_2(\lambda) = \lambda x_1(\lambda)$, then the multiplicity is 3, and a higher multiplicity is not possible. In the latter case one obtains a Jordan chain as $[x_1(0); 0]$, $[x_1'(0); x_1(0)]$, $[x_1''(0); 2x_1'(0)]$.
2. For the quadratic eigenvalue problem in (115.2), $\det T(\lambda) = \lambda^4 - \lambda^3 - 3\lambda^2 + \lambda + 2$. Hence, $\hat{\lambda} = -1$ is an eigenvalue with algebraic multiplicity 2 and geometric multiplicity 1. From

$$T(-1)\mathbf{x}_0 = \begin{bmatrix} -6 & 6 \\ -12 & 12 \end{bmatrix} \mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \; T(-1)\mathbf{x}_1 + T'(-1)\mathbf{x}_0 = \begin{bmatrix} -6 & 6 \\ -12 & 12 \end{bmatrix} \mathbf{x}_1 + \begin{bmatrix} 5 & -5 \\ 10 & -10 \end{bmatrix} \mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

   it follows that $\mathbf{x} = [1; 1]$ is an eigenvector corresponding to $\hat{\lambda}$, and a generalized eigenvector as well. Then for $X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $J = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$ the pair (X, J) is a regular pair of order 2, namely the Jordan pair corresponding to $\hat{\lambda} = -1$.

115.3 Variational Characterization of Eigenvalues

Variational characterizations of eigenvalues are very powerful tools when studying self-adjoint linear operators on a Hilbert space. Many things can be easily proved using these characterizations; for example, bounds for eigenvalues, comparison theorems, interlacing results and monotonicity of eigenvalues, to name just a few.

This section presents similar results for nonlinear eigenvalue problems. A minmax characterization was first proved in [Duf55] for overdamped quadratic eigenproblems, generalized in [Rog64] to general overdamped, and in [VW82] to non-overdamped problems. Although the characterizations also hold for infinite dimensional problems [Had68, VW82] the presentation here is restricted to the finite dimensional case.

We assume in this whole section that $J \subset \mathbb{R}$ is an open interval (which may be unbounded), and we consider a family of Hermitian matrices $T : J \to \mathbb{C}^{n\times n}$ depending continuously on the parameter $\lambda \in J$, such that the following two conditions are satisfied

(i) For every $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x} \neq 0$ the real equation

$$f(\lambda; \mathbf{x}) := \mathbf{x}^* T(\lambda)\mathbf{x} = 0 \tag{115.4}$$

has at most one solution $\lambda =: p(\mathbf{x})$ in J. Then (115.4) implicitly defines a (nonlinear) functional on some domain $D(p)$.

(ii)

$$(\lambda - p(\mathbf{x}))f(\lambda; \mathbf{x}) > 0 \quad \text{for every } \mathbf{x} \in D(p) \text{ and every } \lambda \in J,\ \lambda \neq p(\mathbf{x}). \tag{115.5}$$

Definitions:

The functional $p : D(p) \to J$ is called the **Rayleigh functional**.

If $D(p) = \mathbb{C}^n \setminus \{0\}$, then the problem $T(\lambda)\mathbf{x} = 0$ is called **overdamped**.

An eigenvalue $\hat{\lambda} \in J$ of $T(\cdot)$ is a j**th eigenvalue** if $\mu = 0$ is the j largest eigenvalue of the matrix $T(\hat{\lambda})$.

Facts:

In this subsection we denote by S_j the set of all j dimensional subspaces of $\mathbb{C}^n$. The following facts for which no specific reference is given can be found in [Had68, VW82, Vos09].

1. $D(p)$ is an open set in $\mathbb{C}^n$.
2. $p(\alpha\mathbf{x}) = p(\mathbf{x})$ for every $\mathbf{x} \in D(p)$ and every $\alpha \in \mathbb{C} \setminus \{0\}$.
3. If $T(\cdot)$ is differentiable in a neighborhood of an eigenvalue $\hat{\lambda}$ and $\hat{\mathbf{x}}^* T'(\hat{\lambda})\hat{\mathbf{x}} \neq 0$ for a corresponding eigenvector $\hat{\mathbf{x}}$, then $\hat{\mathbf{x}}$ is a stationary point of p, i.e. $|p(\hat{\mathbf{x}} + \mathbf{h}) - p(\hat{\mathbf{x}})| = o(\|\mathbf{h}\|)$. In the real case $T : J \to \mathbb{R}^{n\times n}$, $J \subset \mathbb{R}$, we have $\nabla p(\hat{\mathbf{x}}) = 0$.
4. For every $j \in \{1, \ldots, n\}$ there is at most one jth eigenvalue of $T(\cdot)$.
5. $T(\cdot)$ has at most n eigenvalues in J.
6. [Rog64] If $T(\cdot)$ is overdamped, then $T(\cdot)$ has exactly n eigenvalues in J.
7. If
$$\lambda_j := \inf_{V \in S_j,\ V \cap D(p) \neq \emptyset} \ \sup_{\mathbf{x} \in V \cap D(p)} p(\mathbf{x}) \in J,$$
then λ_j is a jth eigenvalue of $T(\cdot)$.
8. (minmax characterization) If $\lambda_j \in J$ is a jth eigenvalue of $T(\cdot)$, then
$$\lambda_j := \min_{V \in S_j,\ V \cap D(p) \neq \emptyset} \ \max_{\mathbf{x} \in V \cap D(p)} p(\mathbf{x}) \in J.$$
The minimum is attained for an invariant subspace of the matrix $T(\lambda_j)$ corresponding to its j largest eigenvalues. The maximum is attained for some $\mathbf{x} \in \ker T(\lambda_j)$.

9. Let $\lambda_1 := \inf_{\mathbf{x}\in D(p)} p(\mathbf{x}) \in J$ and $\lambda_j \in J$ for some $j \in \{1,\dots,n\}$. Then for every $k \in \{1,\dots,j\}$ there exists $U_k \in S_k$ with $U_k \subset D(p)\cup\{0\}$ and $\lambda_k := \max_{\mathbf{x}\in U_k,\ \mathbf{x}\neq 0}\ p(\mathbf{x})$. Hence,
$$\lambda_k := \min_{V\in S_k,\ V\subset D(p)\cup\{0\}}\ \max_{\mathbf{x}\in V,\ \mathbf{x}\neq 0}\ p(\mathbf{x}) \in J \quad \text{for } k = 1,\dots,j.$$
10. [Vos03] (maxmin characterization) Assume that there exists a jth eigenvalue $\lambda_j \in J$. Then
$$\lambda_j = \max_{V\in S_{j-1},\ V^\perp\cap D(p)\neq\emptyset}\ \inf_{\mathbf{x}\in V^\perp\cap D(p)}\ p(\mathbf{x}).$$
The maximum is attained for every invariant subspace of $T(\lambda_j)$ corresponding to its $j-1$ largest eigenvalues.
11. Let $\lambda_j \in J$ be a jth eigenvalue of $T(\cdot)$ and $\lambda \in J$. Then
$$\lambda \left\{\begin{matrix}<\\=\\>\end{matrix}\right\} \lambda_j \quad \Longleftrightarrow \quad \max_{V\in S_j}\ \min_{\mathbf{x}\in V,\mathbf{x}\neq 0} \frac{\mathbf{x}^*T(\lambda)\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \left\{\begin{matrix}<\\=\\>\end{matrix}\right\} 0.$$
12. (orthogonality) [Had68] Let $T(\cdot)$ be differentiable in J. Then eigenvectors can be chosen orthogonal with respect to the generalized scalar product
$$[\mathbf{x},\mathbf{y}] := \begin{cases} \mathbf{y}^*\dfrac{T(p(\mathbf{x})) - T(p(\mathbf{y}))}{p(\mathbf{x}) - p(\mathbf{y})}\mathbf{x}, & \text{if } p(\mathbf{x}) \neq p(\mathbf{y}) \\ \mathbf{y}^*T'(p(\mathbf{x}))\mathbf{x}, & \text{if } p(\mathbf{x}) = p(\mathbf{y}) \end{cases}$$
which is symmetric and homogeneous, but in general is not bilinear.
If $T(\cdot)$ is differentiable und condition (ii) strengthened to $x^*T'(p(x))x > 0$ for every $x \in D$, then $[\cdot,\cdot]$ is definite, i.e. $[\mathbf{x},\mathbf{x}] > 0$ for every $x \in D(p)$.
13. (Rayleigh's principle) Assume that J contains $\lambda_1,\dots,\lambda_{j-1}$ where λ_k is a kth eigenvalue of $T(\cdot)$, and let $\mathbf{x}_k$, $k = 1,\dots,j-1$ be a corresponding eigenvectors. If
$$\lambda_j := \inf\{p(\mathbf{x})\ :\ \mathbf{x}\in D(p),\ [\mathbf{x},\mathbf{x}_k] = 0,\ k = 1,\dots,j-1\} \in J,$$
then λ_j is a jth eigenvalue of $T(\cdot)$.
14. (Sylvester's law; overdamped case) Assume that $T(\cdot)$ is overdamped. For $\sigma \in J$ let (π,ν,δ) be the inertia of $T(\sigma)$. Then $T(\cdot)$ has π eigenvalues that are smaller than σ, ν eigenvalues that exceed σ, and if $\delta \neq 0$, then σ is an eigenvalue of multiplicity δ.
15. (Sylvester's law; extreme eigenvalues) Assume that $T(\mu)$ is negative definite for some $\mu \in J$, and for $\sigma > \mu$ let (π,ν,δ) be the inertia of $T(\sigma)$. Then $T(\cdot)$ has exactly π eigenvalues in J that are smaller than σ.
16. (Sylvester's law; general case) Let $\mu \in J$, and assume that for every r dimensional subspace V with $V\cap D(p) \neq \emptyset$ there exists $\mathbf{x} \in V\cap D(p)$ with $p(\mathbf{x}) > \mu$. For $\sigma \in J$, $\sigma > \mu$ let (π,ν,δ) be the inertia of $T(\sigma)$. Then for $j = r,\dots,\pi$ there exists a jth eigenvalue λ_j of $T(\cdot)$ in $[\mu,\sigma)$.

Examples:

1. [Duf55] The quadratic pencil $Q(\lambda) := \lambda^2 A + \lambda B + C$ with positive definite $A,B,C \in \mathbb{C}^{n\times n}$ is overdamped if and only if $d(\mathbf{x}) := (\mathbf{x}^*B\mathbf{x})^2 - 4(\mathbf{x}^*A\mathbf{x})(\mathbf{x}^*C\mathbf{x}) > 0$ for every $\mathbf{x} \in \mathbb{C}^n\setminus\{0\}$. For $\mathbf{x} \neq 0$ the quadratic equation $\mathbf{x}^*Q(\lambda)\mathbf{x} = 0$ has two real solutions $p_\pm(\mathbf{x}) = (-\mathbf{x}^*B\mathbf{x} \pm \sqrt{d(\mathbf{x})})/(2\mathbf{x}^*A\mathbf{x})$, and $\gamma_- := \sup_{\mathbf{x}\neq 0} p_-(\mathbf{x}) < \gamma_+ := \inf_{\mathbf{x}\neq 0} p_+(\mathbf{x})$.
$Q(\cdot)$ has n eigenvalues in $(-\infty,\gamma_+)$ which are minmax values of p_- and n eigenvalues in $(\gamma_-,0)$ which are minmax values of p_+.

2. Assume that $Q(\cdot)$ as in the last example is not necessarily overdamped, and let $\mathrm{in}(Q(\sigma)) = (\pi, \nu, \delta)$ denote the inertia of $Q(\sigma)$. If $\sigma < \gamma_+ := \inf_{\mathbf{x}\neq 0}\{p_+(\mathbf{x}) : p_+(\mathbf{x}) \in \mathbb{R}\}$, then $Q(\cdot)$ has exactly ν eigenvalues in $(-\infty, \sigma)$, and if $\sigma > \gamma_- := \sup_{\mathbf{x}\neq 0}\{p_-(\mathbf{x}) : p_-(\mathbf{x}) \in \mathbb{R}\}$, then $Q(\cdot)$ has ν eigenvalues in $(\sigma, 0)$.
   If $\mu_{\min}$ and $\mu_{\max}$ are the minimal and maximal eigenvalues of $C\mathbf{x} = \mu A\mathbf{x}$, then $-\sqrt{\mu_{\max}} \leq \gamma_+$ and $-\sqrt{\mu_{\min}} \geq \gamma_-$.
   If $\kappa_{\min}$ and $\kappa_{\max}$ are the minimal and maximal eigenvalues of $C\mathbf{x} = \kappa B\mathbf{x}$, respectively, then $-2\kappa_{\max} \leq \gamma_+$ and $-2\kappa_{\min} \geq \gamma_-$.

115.4 General Rayleigh Functionals

Whereas Section 115.3 presupposes the existence and uniqueness of a Rayleigh functional for problems allowing for a variational characterization, this section extends its definition to more general eigenproblems. It collects results on the existence and approximation properties of a Rayleigh functional in a vicinity of eigenvectors corresponding to algebraically simple eigenvalues. The material in this section is mostly taken from [Sch08, SS10].

Definitions:

Let $T : D \to \mathbb{C}^{n\times n}$ be a matrix valued mapping, which is analytic, or which is differentiable with Lipschitz continuous derivative in the real case.

Let $(\hat\lambda, \hat{\mathbf{x}})$ be an eigenpair of $T(\cdot)$, and define neighborhoods $B(\hat\lambda, \tau) := \{\lambda \in \mathbb{C} : |\lambda - \hat\lambda| < \tau\}$ and $\mathcal{K}_\varepsilon(\hat{\mathbf{x}}) := \{\mathbf{x} \in \mathbb{C}^n : \angle(\mathrm{Span}\{\mathbf{x}\}, \mathrm{Span}\{\hat{\mathbf{x}}\}) \leq \varepsilon\}$ of $\hat\lambda$ and $\hat{\mathbf{x}}$, respectively.

$p : \mathcal{K}_\varepsilon \to B(\hat\lambda, \tau)$ is a **(one-sided) Rayleigh functional** if the following conditions hold:

(i) $p(\alpha\mathbf{x}) = p(\mathbf{x})$ for every $\alpha \in \mathbb{C}$, $\alpha \neq 0$
(ii) $\mathbf{x}^*T(p(\mathbf{x}))\mathbf{x} = 0$ for every $\mathbf{x} \in \mathcal{K}_\varepsilon(\hat{\mathbf{x}})$
(iii) $\mathbf{x}^*T'(p(\mathbf{x}))\mathbf{x} \neq 0$ for every $\mathbf{x} \in \mathcal{K}_\varepsilon(\hat{\mathbf{x}})$.

Let $(\hat{\mathbf{y}}, \hat\lambda, \hat{\mathbf{x}})$ be an eigentriplet of $T(\cdot)$.

$p : \mathcal{K}_\varepsilon(\hat{\mathbf{x}}) \times \mathcal{K}_\varepsilon(\hat{\mathbf{y}}) \to B(\hat\lambda, \tau)$ is a **two-sided Rayleigh functional** if the following conditions hold for every $\mathbf{x} \in \mathcal{K}_\varepsilon(\hat{\mathbf{x}})$ and $\mathbf{y} \in \mathcal{K}_\varepsilon(\hat{\mathbf{y}})$:

(i) $p(\alpha\mathbf{x}, \beta\mathbf{y}) = p(\mathbf{x}, \mathbf{y})$ for every $\alpha, \beta \in \mathbb{C} \setminus \{0\}$,
(ii) $\mathbf{y}^*T(p(\mathbf{x}, \mathbf{y}))\mathbf{x} = 0$,
(iii) $\mathbf{y}^*T'(p(\mathbf{x}, \mathbf{y}))\mathbf{x} \neq 0$.

The **generalized Rayleigh quotient** (which was introduced in [Lan61] only for polynomial eigenvalue problems) is defined as

$$p_L : \mathcal{K}_\varepsilon(\hat{\mathbf{x}}) \times B(\hat\lambda, \tau) \times \mathcal{K}_\varepsilon(\hat{\mathbf{y}}) \to B(\hat\lambda, \tau),\ p_L(\mathbf{y}, \lambda, \mathbf{x}) := \lambda - \frac{\mathbf{y}^*T(\lambda)\mathbf{x}}{\mathbf{y}^*T'(\lambda)\mathbf{x}}.$$

Facts:

The following facts can be found in [Sch08, SS10].

1. Let $(\hat{\mathbf{y}}, \hat\lambda, \hat{\mathbf{x}})$ be an eigentriplet of $T(\cdot)$ with $\|\hat{\mathbf{x}}\| = \|\hat{\mathbf{y}}\| = 1$, and assume that $\hat{\mathbf{y}}^*T'(\hat\lambda)\hat{\mathbf{x}} \neq 0$. Then there exist $\varepsilon > 0$ and $\tau > 0$ such that the two-sided Rayleigh functional is defined in $\mathcal{K}_\varepsilon(\hat x) \times \mathcal{K}_\varepsilon(\hat{\mathbf{y}})$, and

$$|p(\mathbf{x}, \mathbf{y}) - \hat\lambda| \leq \frac{8}{3}\frac{\|T(\hat\lambda)\|}{|\hat{\mathbf{y}}^*T'(\hat\lambda)\hat{\mathbf{x}}|}\tan\xi\tan\eta,$$

where $\xi := \angle(\mathrm{Span}\{\mathbf{x}\}, \mathrm{Span}\{\hat{\mathbf{x}}\})$ and $\eta := \angle(\mathrm{Span}\{\mathbf{y}\}, \mathrm{Span}\{\hat{\mathbf{y}}\})$.

2. Under the conditions of Fact 1 let $\xi < \pi/3$ and $\eta < \pi/3$. Then
$$|p(\mathbf{x},\mathbf{y}) - \hat{\lambda}| \leq \frac{32}{3}\frac{\|T(\hat{\lambda})\|}{|\hat{\mathbf{y}}^* T'(\hat{\lambda})\hat{\mathbf{x}}|}\|\mathbf{x}-\hat{\mathbf{x}}\|\|\mathbf{y}-\hat{\mathbf{y}}\|.$$
3. Under the conditions of Fact 1 the two-sided Rayleigh functional is stationary at $(\hat{\mathbf{x}},\hat{\mathbf{y}})$, i.e. $|p(\hat{\mathbf{x}}+\mathbf{s},\hat{\mathbf{y}}+\mathbf{t}) - \hat{\lambda}| = O((\|\mathbf{s}\| + \|\mathbf{t}\|)^{\mathbf{2}})$.
4. Let $(\hat{\lambda},\hat{\mathbf{x}})$ be an eigenpair of $T(\cdot)$ with $\|\hat{\mathbf{x}}\| = 1$ and $\hat{\mathbf{x}}^* T'(\hat{\lambda})\hat{\mathbf{x}} \neq 0$, and suppose that $T(\hat{\lambda}) = T(\hat{\lambda})^*$. Then there exist $\varepsilon > 0$ and $\tau > 0$ such that the one-sided Rayleigh functional $p(\cdot)$ is defined in $\mathcal{K}_\varepsilon(\hat{\mathbf{x}})$, and
$$|p(\mathbf{x}) - \hat{\lambda}| \leq \frac{8}{3}\frac{\|T(\hat{\lambda})\|}{|\hat{\mathbf{x}}^* T'(\hat{\lambda})\hat{\mathbf{x}}|}\tan^2\xi,$$
where $\xi := \angle(\mathrm{Span}\{\mathbf{x}\}, \mathrm{Span}\{\hat{\mathbf{x}}\})$.
5. Let $(\hat{\lambda},\hat{\mathbf{x}})$ be an eigenpair of $T(\cdot)$ with $\|\hat{\mathbf{x}}\| = 1$ and $\hat{\mathbf{x}}^* T'(\hat{\lambda})\hat{\mathbf{x}} \neq 0$. Then there exist $\varepsilon > 0$ and $\tau > 0$ such that the one-sided Rayleigh functional $p(\cdot)$ is defined in $\mathcal{K}_\varepsilon(\hat{\mathbf{x}})$, and
$$|p(\mathbf{x}) - \hat{\lambda}| \leq \frac{10}{3}\frac{\|T(\hat{\lambda})\|}{|\hat{\mathbf{x}}^* T'(\hat{\lambda})\hat{\mathbf{x}}|}\tan\xi,$$
where $\xi := \angle(\mathrm{Span}\{\mathbf{x}\}, \mathrm{Span}\{\hat{\mathbf{x}}\})$.
6. Let $\hat{\mathbf{x}}$ be a right eigenvector of $T(\cdot)$ corresponding to $\hat{\lambda}$, and $\hat{\mathbf{x}}^* T'(\hat{\lambda})\hat{\mathbf{x}} \neq 0$. The one-sided Rayleigh functional p is only stationary at $\hat{\mathbf{x}}$ if $\hat{\mathbf{x}}$ is also a left eigenvector.
7. The generalized Rayleigh quotient p_L is obtained when applying Newton's method to the equation defining the two-sided Rayleigh functional for fixed $\mathbf{x}$ and $\mathbf{y}$.
8. [Lan61] Let $(\hat{\mathbf{y}},\hat{\lambda},\hat{\mathbf{x}})$ be an eigentriplet of $T(\cdot)$ with $\hat{\mathbf{y}}^* T'(\hat{\lambda})\hat{\mathbf{x}} \neq 0$. Then the generalized Rayleigh quotient p_L is stationary at $(\hat{\mathbf{y}},\hat{\lambda},\hat{\mathbf{x}})$.
9. Under the conditions of Fact 1 the generalized Rayleigh quotient p_L is defined for all $\lambda \in B(\hat{\lambda},\tau)$ and $(\mathbf{x},\mathbf{y}) \in \mathcal{K}_\varepsilon(\hat{\mathbf{x}}) \times \mathcal{K}_\varepsilon(\hat{\mathbf{y}})$, and
$$|p_L(\mathbf{y},\lambda,\mathbf{x}) - \hat{\lambda}| \leq \frac{4\|T(\hat{\lambda})\|}{|\hat{\mathbf{y}}^* T'(\hat{\lambda})\hat{\mathbf{x}}|}\tan\xi\tan\eta + \frac{2L}{|\hat{\mathbf{y}}^* T'(\hat{\lambda})\hat{\mathbf{x}}|}\frac{|\lambda-\hat{\lambda}|^2}{\cos\xi\cos\eta},$$
where L denotes the Lipschitz constant of $T'(\cdot)$.

115.5 Methods for dense eigenvalue problems

The size of the eigenvalue problems that can be treated with the numerical methods considered in this section is limited to a few thousands depending on the available storage capacities. Moreover, they require several factorizations of varying matrices to approximate one eigenvalue, and therefore, they are not appropriate for large and sparse problems. However, they are needed to solve the projected eigenproblem in most of the iterative projection methods for sparse problems.

For general nonlinear eigenvalue problems, the classical approach is to formulate the eigenvalue problem as a system of nonlinear equations and to use variations of Newton's method or the inverse iteration method. Thus, these methods are local and therefore not guaranteed to converge, but as for linear eigenvalue problems their basin of convergence can be enlarged using homotopy methods [DP01] or trust region strategies [YMW07].

Facts:

1. [Kub70] Let $T(\lambda)P(\lambda) = Q(\lambda)R(\lambda)$ be the QR factorization of $T(\lambda)$, where $P(\lambda)$ is a permutation matrix which is chosen such that the diagonal elements $r_{jj}(\lambda)$ of $R(\lambda)$ are decreasing in magnitude, i.e. $|r_{11}(\lambda)| \geq |r_{22}(\lambda)| \geq \cdots \geq |r_{nn}(\lambda)|$. Then λ is an eigenvalue of $T(\cdot)$ if and only if $r_{nn}(\lambda) = 0$.
   Applying Newton's method to this equation, one obtains the iteration
   $$\lambda_{k+1} = \lambda_k - \frac{1}{\mathbf{e}_n^T Q(\lambda_k)^* T'(\lambda_k) P(\lambda_k) R(\lambda_k)^{-1} \mathbf{e}_n}$$
   for approximations to an eigenvalue of problem $T(\lambda)x = 0$, where $\mathbf{e}_n$ denotes the nth unit vector.
   Approximations to left and right eigenvectors can be obtained from $\mathbf{y}_k = Q(\lambda_k)\mathbf{e}_n$ and $\mathbf{x}_k = P(\lambda_k)R(\lambda_k)^{-1}\mathbf{e}_n$.
   However, this relatively simple idea is not efficient, since it computes eigenvalues one at a time and needs several $O(n^3)$ factorizations per eigenvalue. It is, however, useful in the context of iterative refinement of computed eigenvalues and eigenvectors.
2. [AR68] Applying Newton's method to the nonlinear system
   $$F(\mathbf{x}, \lambda) := \begin{pmatrix} T(\lambda)\mathbf{x} \\ \mathbf{v}^*\mathbf{x} - 1 \end{pmatrix} = 0$$
   where $\mathbf{v} \in \mathbb{C}^n$ is suitably chosen one obtains the *inverse iteration* given in Algorithm 1. Being a variant of Newton's method it converges locally and quadratically for simple eigenpairs.

   Algorithm 1: Inverse iteration
   Require: Initial pair $(\lambda_0, \mathbf{x}_0)$ and normalization vector $\mathbf{v}$ with $\mathbf{v}^*\mathbf{x}_0 = 1$
   1: **for** $k = 0, 1, 2, \ldots$ until convergence **do**
   2: solve $T(\lambda_k)\mathbf{u}_{k+1} = T'(\lambda_k)\mathbf{x}_k$ for $\mathbf{u}_{k+1}$
   3: $\lambda_{k+1} \leftarrow \lambda_k - (\mathbf{v}^*\mathbf{x}_k)/(\mathbf{v}^*\mathbf{u}_{k+1})$
   4: normalize $\mathbf{x}_{k+1} \leftarrow \mathbf{u}_{k+1}/\mathbf{v}^*\mathbf{u}_{k+1}$
   5: **end for**
3. If $T(\cdot)$ is Hermitian such that the general conditions of Section 115.3 are satisfied, one obtains the *Rayleigh functional iteration* if the update of λ_{k+1} in step 3 of Algorithm 1 is replaced with $\lambda_{k+1} \leftarrow p(\mathbf{u}_{k+1})$. This method converges locally and cubically [Rot89] for simple eigenpairs.
4. [Lan61, Ruh73] Replacing the vector $\mathbf{v}$ in the normalization step of inverse iteration for a general matrix function $T(\cdot)$ with $\mathbf{v}_k = T(\lambda_k)^*\mathbf{y}_k$, where $\mathbf{y}_k$ is an approximation to a left eigenvector, the update for λ becomes
   $$\lambda_{k+1} \leftarrow \lambda_k - \frac{\mathbf{y}_k^* T(\lambda_k)\mathbf{x}_k}{\mathbf{y}_k^* T'(\lambda_k)\mathbf{x}_k},$$
   which is the generalized Rayleigh quotient p_L.
5. [Sch08] For general $T(\cdot)$ and simple eigentriplets $(\hat{\mathbf{y}}, \hat{\lambda}, \hat{\mathbf{x}})$ cubic convergence is also achieved by the *two-sided Rayleigh functional iteration* in Algorithm 2. If the linear system in step 2 is solved by factorizing $T(\lambda_k)$, then taking the conjugate transpose the factorization can be reused for the system in step 3. So, the cost of one iteration step is similar to the one of the one-sided Rayleigh functional iteration.

Algorithm 2: Two-sided Rayleigh functional iteration

Require: Initial triplet $(\mathbf{y}_0, \lambda_0, \mathbf{x}_0)$ where $\mathbf{x}_0^*\mathbf{x}_0 = \mathbf{y}_0^*\mathbf{y}_0 = 1$

1: **for** $k = 0, 1, 2, \ldots$ until convergence **do**
2: solve $T(\lambda_k)\mathbf{u}_{k+1} = T'(\lambda_k)\mathbf{x}_k$ for $\mathbf{u}_{k+1}$; $\mathbf{x}_{k+1} \leftarrow \mathbf{u}_{k+1}/\|\mathbf{u}_{k+1}\|$
3: solve $T(\lambda_k)^*\mathbf{v}_{k+1} = T'(\lambda_k)^*\mathbf{y}_k$ for $\mathbf{v}_{k+1}$; $\mathbf{y}_{k+1} \leftarrow \mathbf{v}_{k+1}/\|\mathbf{v}_{k+1}\|$
4: solve $\mathbf{y}_{k+1}^* T(\lambda_{k+1})\mathbf{x}_{k+1} = 0$ for λ_{k+1}
5: **end for**

6. [Neu85] The cost for solving a linear system in each iteration step with a varying matrix is avoided in the *residual inverse iteration* in Algorithm 3 where the matrix $T(\lambda_0)$ is fixed during the whole iteration (or at least for several steps).

Algorithm 3: Residual inverse iteration

Require: Initial pair $(\lambda_0, \mathbf{x}_0)$ and normalization vector $\mathbf{w}$ with $\mathbf{w}^*\mathbf{x}_0 = 1$

1: **for** $k = 0, 1, 2, \ldots$ until convergence **do**
2: solve $\mathbf{w}^* T(\lambda_0)^{-1} T(\lambda_{k+1})\mathbf{x}_k = 0$ for λ_{k+1}
3: solve $T(\lambda_0)\mathbf{u}_k = T(\lambda_{k+1})\mathbf{x}_k$ for $\mathbf{u}_k$
4: set $\mathbf{v}_{k+1} \leftarrow \mathbf{x}_k - \mathbf{u}_k$ and normalize $\mathbf{x}_{k+1} \leftarrow \mathbf{v}_{k+1}/\mathbf{w}^*\mathbf{v}_{k+1}$
5: **end for**

If $T(\cdot)$ is Hermitian and $\hat{\lambda} \in \mathbb{R}$, then the convergence can be improved by determining λ_{k+1} in step 1 via the Rayleigh functional, i.e. solving $\mathbf{x}_k^* T(\lambda_{k+1})\mathbf{x}_k = 0$ for λ_{k+1}. If $T(\cdot)$ is twice continuously differentiable and $\hat{\lambda}$ is algebraically simple, then the residual inverse iteration converges for all $(\lambda_0, \mathbf{x}_0)$ sufficiently close to $(\hat{\lambda}, \hat{\mathbf{x}})$, and

$$\frac{\|\mathbf{x}_{k+1} - \hat{\mathbf{x}}\|}{\|\mathbf{x}_k - \hat{\mathbf{x}}\|} = O(|\lambda_0 - \hat{\lambda}|) \quad \text{and} \quad |\lambda_{k+1} - \hat{\lambda}| = O(\|\mathbf{x}_k - \hat{\mathbf{x}}\|^t)$$

where $t = 2$ in the Hermitian case if λ_{k+1} is updated via the Rayleigh functional, and $t = 1$ in the general case.

7. [Ruh73] The first order approximation $T(\lambda + \sigma)\mathbf{x} = T(\lambda)\mathbf{x} + \sigma T'(\lambda)\mathbf{x} + o(|\sigma|)$ suggests the *method of successive linear problems* in Algorithm 4, which also converges quadratically for simple eigenvalues.

Algorithm 4: Method of successive linear problems

Require: Initial approximation λ_0

1: **for** $k = 0, 1, \ldots$ until convergence **do**
2: solve the linear eigenproblem $T(\lambda_k)\mathbf{u} = \theta T'(\lambda_k)\mathbf{u}$
3: choose an eigenvalue θ smallest in modulus
4: $\lambda_{k+1} = \lambda_k - \theta$
5: **end for**

If $\hat{\lambda}$ is a semi-simple eigenvalue, $\mathbf{x}_k$ converges to a right eigenvector $\hat{\mathbf{x}}$. If $\hat{\mathbf{y}}$ is a left eigenvector corresponding to $\hat{\lambda}$ such that $\hat{\mathbf{y}}^* T'(\hat{\lambda})\hat{\mathbf{x}} \neq 0$ (which is guaranteed for a simple eigenvalue), then the convergence factor is given by (cf. [Jar12])

$$c := \lim_{k\to\infty} \frac{\lambda_{k+1} - \hat{\lambda}}{(\lambda_k - \hat{\lambda})^2} = \frac{1}{2}\frac{\hat{\mathbf{y}}^* T''(\hat{\lambda})\hat{\mathbf{x}}}{\hat{\mathbf{y}}^* T'(\hat{\lambda})\hat{\mathbf{x}}}.$$

8. [Wer70] If the nonlinear eigenvalue problem allows for a variational characterization of its eigenvalues, then the *safeguarded iteration*, which aims at a particular eigenvalue, is a natural choice.

Algorithm 5: Safeguarded iteration for determining an mth eigenvalue
Require: Approximation λ_0 to an mth eigenvalue
1: **for** $k = 0, 1, \ldots$ until convergence **do**
2: determine an eigenvector $\mathbf{x}_k$ corresponding to the m-largest eigenvalue of $T(\lambda_k)$
3: solve $\mathbf{x}_k^* T(\lambda_{k+1})\mathbf{x}_k = 0$ for λ_{k+1}
4: **end for**

Under the conditions of Section 115.3, the safeguarded iteration has the following properties [NV10]:

(i) If $\hat{\lambda}_1 := \inf_{\mathbf{x}\in D(p)} p(\mathbf{x}) \in J$ and $\mathbf{x}_0 \in D$, then the safeguarded iteration with $m = 1$ converges globally to $\hat{\lambda}_1$.

(ii) If $T(\cdot)$ is continuously differentiable and $\hat{\lambda}_m$ is a simple eigenvalue, then the safeguarded iteration converges locally and quadratically to $\hat{\lambda}_m$.

(iii) Let $T(\cdot)$ be twice continuously differentiable and $T'(\hat{\lambda}_m)$ be positive definite. If $\mathbf{x}_k$ in step 3 is chosen to be an eigenvector corresponding to the m largest eigenvalue of the generalized eigenvalue problem $T(\lambda_k)\mathbf{x} = \mu T'(\lambda_k)\mathbf{x}$, then the convergence is even cubic.

9. [SX11] For higher dimensions n it is too costly to solve the occurring linear systems exactly. Szyld and Xue [SX11] studied inexact versions of inverse iteration and residual inverse iteration and proved that the same order of convergence can be achieved as for the exact methods if the respective linear systems are solved sufficiently accurately.

115.6 Iterative projection methods

For sparse linear eigenvalue problems $A\mathbf{x} = \lambda\mathbf{x}$ iterative projection methods like the Lanczos, Arnoldi, rational Krylov or Jacobi–Davidson method are very efficient. Here the dimension of the eigenproblem is reduced by projecting it to a subspace of much smaller dimension, and the reduced problem is solved by a fast technique for dense problems. The subspaces are expanded in the course of the algorithm in an iterative way with the aim that some of the eigenvalues of the reduced matrix become good approximations to some of the wanted eigenvalues of the given large matrix.

Two types of iterative projection methods are in use: methods which expand the subspaces independently of the eigenpair of the projected problem and which take advantage of a normal form of A like the Arnoldi, Lanczos, and rational Krylov method, and methods which aim at a particular eigenpair and choose the expansion such that it has a high approximation potential for a wanted eigenvector like the Jacobi–Davidson method.

For general nonlinear eigenproblems a normal form does not exist. Therefore, generalizations of iterative projection methods to general nonlinear eigenproblems always have to be of the second type. There are essentially two types of these methods, the Jacobi–Davidson method (and its two–sided version) which is based on inverse iteration and the nonlinear Arnoldi method which is based on residual inverse iteration.

Jacobi–Davidson method
Assume that we are given a search space $\mathcal{V}$ and a matrix V with orthonormal columns

containing a basis of $\mathcal{V}$. Let $(\mathbf{y}, \theta)$ be an eigenpair of the projected problem $V^*T(\lambda)V\mathbf{y} = 0$ and $\mathbf{x} = V\mathbf{y}$ be the corresponding Ritz vector. A direction with high approximation potential is given by inverse iteration $\mathbf{v} = T(\theta)^{-1}T'(\theta)\mathbf{x}$, however replacing $\mathbf{v}$ with an inexact solution of the linear system $T(\theta)\mathbf{v} = T'(\theta)\mathbf{x}$ will spoil the favorable approximation properties of inverse iteration.

Actually, we are not interested in the direction $\mathbf{v}$ but in an expansion of $\mathcal{V}$ which contains $\mathbf{v}$, and for every $\alpha \neq 0$ the vector $\mathbf{t} = \mathbf{x} + \alpha\mathbf{v}$ is as qualified as $\mathbf{v}$. It was shown in [Vos07] that the most robust expansion of this type is obtained if $\mathbf{x}$ and $\mathbf{t} := \mathbf{x} + \alpha\mathbf{v}$ are orthogonal, and it is easily seen that this $\mathbf{t}$ solves the so called correction equation

$$\left(I - \frac{T'(\theta)\mathbf{x}\mathbf{x}^*}{\mathbf{x}^*T'(\theta)\mathbf{x}}\right) T(\theta) \left(I - \frac{\mathbf{x}\mathbf{x}^*}{\mathbf{x}^*\mathbf{x}}\right) \mathbf{t} = T(\theta)\mathbf{x}, \quad \mathbf{t} \perp \mathbf{x}.$$

The resulting iterative projection method is called **Jacobi–Davidson method**, a template of which is given in Algorithm 6.

Algorithm 6: Nonlinear Jacobi–Davidson method

Require: Initial basis V, $V^*V = I$; $m = 1$

1: determine preconditioner $K \approx T(\sigma)^{-1}$, σ close to first wanted eigenvalue
2: **while** $m \leq$ number of wanted eigenvalues **do**
3: compute an approximation θ to the mth wanted eigenvalue and corresponding eigenvector $\mathbf{y}$ of the projected problem $T_V(\theta)\mathbf{y} := V^*T(\theta)V\mathbf{y} = 0$
4: determine the Ritz vector $\mathbf{u} = V\mathbf{y}$ and the residual $\mathbf{r} = T(\theta)\mathbf{u}$
5: **if** $\|\mathbf{r}\|/\|\mathbf{u}\| < \epsilon$ **then**
6: accept approximate eigenpair $(\lambda_m, \mathbf{x}_m) := (\theta, \mathbf{u})$; increase $m \leftarrow m + 1$;
7: reduce search space V if indicated
8: determine new preconditioner $K \approx T(\lambda_m)^{-1}$ if necessary
9: choose approximation $(\theta, \mathbf{u})$ to next eigenpair
10: compute residual $\mathbf{r} = T(\theta)\mathbf{u}$;
11: **end if**
12: Find approximate solution of correction equation

$$(I - \frac{T'(\theta)\mathbf{u}\mathbf{u}^*}{\mathbf{u}^*T'(\theta)\mathbf{u}})T(\theta)(I - \frac{\mathbf{u}\mathbf{u}^*}{\mathbf{u}^*\mathbf{u}})\mathbf{t} = -\mathbf{r}, \quad \mathbf{t} \perp \mathbf{u} \tag{115.6}$$

(by preconditioned Krylov solver, e.g.)
13: orthogonalize $\mathbf{t} = \mathbf{t} - VV^*\mathbf{t}$, $\mathbf{v} = \mathbf{t}/\|\mathbf{t}\|$, and expand subspace $V = [V, \mathbf{v}]$
14: update projected problem
15: **end while**

Facts:

1. The Jacobi–Davidson method was introduced for polynomial eigenproblem in [SBF96] and studied for general nonlinear eigenvalue problems in [BV04, Vos07a].
2. As in the linear case the correction equation (115.6) does not have to be solved exactly to maintain fast convergence, but usually a few steps of a Krylov subspace solver with an appropriate preconditioner suffice to obtain a good expansion direction of the search space.
3. In the correction equation (115.6) the operator $T(\theta)$ is restricted to map the subspace $\mathbf{x}^\perp$ into itself. Hence, if $K^{-1} \approx T(\theta)$ is a preconditioner of $T(\theta)$, then a preconditioner

for an iterative solver of (115.6) should be modified correspondingly to

$$\tilde{K} := (I - \frac{T'(\theta)\mathbf{u}\mathbf{u}^*}{\mathbf{u}^*T'(\theta)\mathbf{u}})K^{-1}(I - \frac{\mathbf{u}\mathbf{u}^*}{\mathbf{u}^*\mathbf{u}}).$$

With left-preconditioning equation (115.6) becomes

$$\tilde{K}^{-1}\tilde{T}(\theta)\mathbf{t} = -\tilde{K}^{-1}\mathbf{r},\ \mathbf{t} \perp \mathbf{u} \text{ where } \tilde{T}(\theta) := (I - \frac{T'(\theta)\mathbf{u}\mathbf{u}^*}{\mathbf{u}^*T'(\theta)\mathbf{u}})T(\theta)(I - \frac{\mathbf{u}\mathbf{u}^*}{\mathbf{u}^*\mathbf{u}}).$$

Taking into account the projectors in the preconditioner, i.e. using $\tilde{K}$ instead of K in a preconditioned Krylov solver, raises the cost only slightly. In every step one has to solve one linear system $K\mathbf{w} = \mathbf{y}$, and to initialize the solver requires only one additional solve.

4. In step 1 of Algorithm 6 any preinformation such as a small number of known approximate eigenvectors of problem (115.1) corresponding to eigenvalues close to σ or of eigenvectors of a contiguous problem can and should be used.
   If no information on eigenvectors is at hand, and if one is interested in eigenvalues close to the parameter $\sigma \in D$, one can choose an initial vector at random, execute a few Arnoldi steps for the linear eigenproblem $T(\sigma)\mathbf{u} = \theta\mathbf{u}$ or $T(\sigma)\mathbf{u} = \theta T'(\sigma)\mathbf{u}$, and choose the eigenvector corresponding to the smallest eigenvalue in modulus or a small number of Schur vectors as initial basis of the search space.
   Starting with a random vector without this preprocessing usually will yield a value λ_m in step 4 which is far away from σ and will avert convergence.
5. As the subspaces expand in the course of the algorithm the increasing storage or the computational cost for solving the projected eigenvalue problems may make it necessary to restart the algorithm and purge some of the basis vectors. Since a restart destroys information on the eigenvectors and particularly on the one the method is just aiming at, the method is restarted only if an eigenvector has just converged.
   An obvious way to restart is to determine a Ritz pair $(\mu, \mathbf{u})$ from the projection to the current search space span(V) approximating an eigenpair wanted next, and to restart the Jacobi–Davidson method with this single vector $\mathbf{u}$. However, this may discard too much valuable information contained in span(V), and may slowdown the speed of convergence too much. Therefore, thick restarts with subspaces spanned by the Ritz vector $\mathbf{u}$ and a small number of eigenvector approximations obtained in previous steps which correspond to eigenvalues closest to μ are preferable.
6. A crucial point in iterative methods for general nonlinear eigenvalue problems when approximating more than one eigenvalue is to inhibit the method from converging to the same eigenvalue repeatedly. For linear eigenvalue problems locking of already converged eigenvectors can be achieved using an incomplete Schur factorization. For nonlinear problems allowing for a variational characterization of its eigenvalues one can determine the eigenpairs one after another solving the projected problem by safeguarded iteration [BV04]. For general nonlinear eigenproblems a locking procedure based on invariant pairs was introduced in [Eff12] (cf. Subsection 7).
7. Often the matrix function $T(\cdot)$ is given in the following form $T(\lambda) := \sum_{j=1}^{m} f_j(\lambda)A_j$ where $f_j : \Omega \to \mathbb{C}$ are continuous functions and $A_j \in \mathbb{C}^{n\times n}$ are fixed matrices. Then the projected problem can be updated easily appending one row and column to each of the projected matrices V^*A_jV.

Two-sided Jacobi–Davidson method

In Algorithm 6 approximations to an eigenvalue are obtained in step 3 from a Galerkin projection of $T(\lambda)\mathbf{x} = 0$ to the search space Span (V) for right eigenvectors. Computing a

left search space also with a correction equation for left eigenvectors and applying a Petrov-Galerkin projection one arrives at the *Two-sided Jacobi-Davidson method* in Algorithm 7 (where only the computation of one eigentriplet is considered):

Algorithm 7: Two-sided Jacobi–Davidson method

Require: Initial bases U with $U^*U = I$ and V with $V^*V = I$

1: **while** not converged **do**
2: solve $V^*T(\theta)U\mathbf{c} = 0$ and $U^*T(\theta)^*V\mathbf{d} = 0$ for $(\theta, \mathbf{c}, \mathbf{d})$
3: determine Ritz vectors $\mathbf{u} = U\mathbf{c}$ and $\mathbf{v} = V\mathbf{d}$ and residuals $\mathbf{r}_u = T(\theta)\mathbf{u}$, $\mathbf{r}_v = T(\theta)^*\mathbf{v}$
4: **if** $\min(\|\mathbf{r_u}\|/\|\mathbf{u}\|, \|\mathbf{r_v}\|/\|\mathbf{v}\|) < \epsilon$ **then**
5: accept approximate eigentriplet $(\mathbf{v}, \theta, \mathbf{u})$; STOP
6: **end if**
7: Solve (approximately) correction equations

$$(I - \frac{T'(\theta)\mathbf{u}\mathbf{v}^*}{\mathbf{v}^*T'(\theta)\mathbf{u}})T(\theta)(I - \frac{\mathbf{u}\mathbf{u}^*}{\mathbf{u}^*\mathbf{u}})\mathbf{s} = -\mathbf{r}_u, \quad \mathbf{s} \perp \mathbf{u},$$

$$(I - \frac{T'(\theta)^*\mathbf{v}\mathbf{u}^*}{\mathbf{u}^*T'(\theta)^*\mathbf{v}})T(\theta)^*(I - \frac{\mathbf{v}\mathbf{v}^*}{\mathbf{v}^*\mathbf{v}})\mathbf{t} = -\mathbf{r}_v, \quad \mathbf{t} \perp \mathbf{v}$$

8: orthogonalize $\mathbf{s} = \mathbf{s} - UU^*\mathbf{s}$, $\mathbf{s} = \mathbf{s}/\|\mathbf{s}\|$, and expand left search space $U = [U, \mathbf{s}]$
9: orthogonalize $\mathbf{t} = \mathbf{t} - VV^*\mathbf{t}$, $\mathbf{t} = \mathbf{t}/\|\mathbf{t}\|$, and expand right search space $V = [V, \mathbf{t}]$
10: **end while**

Facts:

8. [Sch08] θ as computed in step 2 is the value of the two-sided Rayleigh functional at $(\mathbf{u}, \mathbf{v})$, and one therefore may expect local cubic convergence for simple eigenvalues.
9. [HS03] The correction equation in step 7 of Algorithm 7 can be replaced with

$$(I - \frac{T'(\theta)\mathbf{u}\mathbf{v}^*}{\mathbf{v}^*T'(\theta)\mathbf{u}})T(\theta)(I - \frac{T'(\theta)\mathbf{u}\mathbf{v}^*}{\mathbf{v}^*T'(\theta)\mathbf{u}})\mathbf{s} = -\mathbf{r}_u, \quad \mathbf{s} \perp \mathbf{u},$$

$$(I - \frac{T'(\theta)^*\mathbf{v}\mathbf{u}^*}{\mathbf{u}^*T'(\theta)^*\mathbf{v}})T(\theta)^*(I - \frac{T'(\theta)^*\mathbf{v}\mathbf{u}^*}{\mathbf{u}^*T'(\theta)^*\mathbf{v}})\mathbf{t} = -\mathbf{r}_v, \quad \mathbf{t} \perp v.$$

   This variant was suggested in [HS03] for linear eigenvalue problems, and its generalization to the nonlinear problem is obvious. Since again θ is the value of the two-sided Rayleigh functional the convergence should also be cubic.
10. [SS06] Replacing the correction equations with

$$(I - \mathbf{v}\mathbf{v}^*)T(\theta)(I - \mathbf{u}\mathbf{u}^*)\mathbf{s} = -\mathbf{r}_u, \quad \mathbf{s} \perp \mathbf{u},$$

$$(I - \mathbf{u}\mathbf{u}^*)T(\theta)^*(I - \mathbf{v}\mathbf{v}^*)\mathbf{t} = -\mathbf{r}_v, \quad \mathbf{t} \perp \mathbf{v}$$

    one obtains the **primal-dual Jacobi-Davidson method** which was shown to be quadratically convergent.

Nonlinear Arnoldi method

Expanding the current search space $\mathcal{V}$ by the direction $\hat{\mathbf{v}} = \mathbf{x} - T^{-1}(\sigma)T(\theta)\mathbf{x}$ as suggested by residual inverse iteration generates similar robustness problems as for inverse iteration. If $\hat{\mathbf{v}}$ is close to the desired eigenvector, then an inexact evaluation of $\hat{\mathbf{v}}$ spoils the favorable approximation properties of residual inverse iteration.

Similarly as in the Jacobi–Davidson method one could replace $\hat{\mathbf{v}}$ by $\mathbf{z} := \mathbf{x} + \alpha\hat{\mathbf{v}}$ where α is chosen that $\mathbf{x}^*\mathbf{z} = 0$, and one could determine an approximation to $\mathbf{z}$ solving a correction equation. However, since the new search direction is orthonormalized against the previous search space $\mathcal{V}$ and since $\mathbf{x}$ is contained in $\mathcal{V}$ we may choose the new direction $\mathbf{v} = T(\sigma)^{-1}T(\theta)\mathbf{x}$ as well. This direction satisfies the orthogonality condition $\mathbf{x}^*\mathbf{v} = 0$ at least in the limit as θ approaches a simple eigenvalue $\hat{\lambda}$ (cf. [Vos07]), i.e. $\lim_{\theta\to\hat{\lambda}} \mathbf{x}^*T(\sigma)^{-1}T(\theta)\mathbf{x} = 0$.

A template for the preconditioned **nonlinear Arnoldi method** with restarts and varying preconditioner is just like Algorithm 6. Only step 12 has to be replaced with $\mathbf{t} = K\mathbf{r}$.

Facts:

11. The general remarks about the initial approximation to the eigenvector, restarts and locking following the Jacobi–Davidson method apply to the nonlinear Arnoldi method also.
12. Since the residual inverse iteration with fixed pole σ converges (at least) linearly, and the contraction rate satisfies $O(|\sigma-\lambda_m|)$, it is reasonable to update the preconditioner if the convergence (measured by the quotient of the last two residual norms before convergence) has become too slow.
13. The nonlinear Arnoldi method was introduced for quadratic eigenvalue problems in [Mee01] and for general nonlinear eigenvalue problems in [Vos04].
14. [LBL10] studies a variant that avoids complex arithmetic augmenting the search space by two vectors, the real and imaginary part of the expansion $\mathbf{t} = K\mathbf{r}$.

115.7 Methods using invariant pairs

One of the most important problems when determining more than one eigenpair of a nonlinear eigenvalue problem is to prevent the method from determining the same pair repeatedly. Jordan chains are conceptually elegant but unstable under perturbations. More robust concepts for computing several eigenvalues along with the corresponding (generalized) eigenvectors were introduced only recently and are based on invariant pairs [Kre09, BT09].

It is convenient to consider the nonlinear eigenvalue problem in the following form:

$$T(\lambda)\mathbf{x} := \sum_{j=1}^{m} f_j(\lambda)A_j\mathbf{x} = 0 \tag{115.7}$$

where $f_j : \Omega \to \mathbb{C}$ are analytic functions and $A_j \in \mathbb{C}^{n\times n}$ are fixed matrices.

Definitions:

Let the eigenvalues of $S \in \mathbb{C}^{k\times k}$ be contained in Ω and let $X \in \mathbb{C}^{n\times k}$. Then (X, S) is called **invariant pair** of the nonlinear eigenvalue problem (115.7) if

$$\sum_{j=1}^{m} A_j X f_j(S) = 0.$$

A pair $(X, S) \in \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k}$ is **minimal** if there is $\ell \in \mathbb{N}$ such that the matrix

$$V_\ell(X, S) := \begin{bmatrix} X \\ XS \\ \vdots \\ XS^{\ell-1} \end{bmatrix}$$

has rank k.

The smallest such ℓ is called the **minimality index** of (X, S).

An invariant pair (X, S) is called **simple** if (X, S) is minimal and the algebraic multiplicities of the eigenvalues of S are identical to the ones of the corresponding eigenvalues of $T(\cdot)$.

Facts:

The following facts for which no specific reference is given can be found in [Kre09].

1. Let (X, S) be a minimal invariant pair of (115.7). Then the eigenvalues of S are eigenvalues of $T(\cdot)$.
2. By the Cayley–Hamilton theorem the minimality index of a minimal pair can not exceed k.
3. [BK11] For a regular matrix polynomial of degree m the minimality index of a minimal invariant pair can not exceed m.
4. [Eff12] Let $p_0, \ldots, p_{\ell-1}$ be a basis for the polynomials of degree less than ℓ. Then the pair (X, S) is minimal with minimality index at most ℓ if and only if
$$V_\ell^p(X, S) = \begin{bmatrix} Xp_0(S) \\ \vdots \\ Xp_{\ell-1}(S) \end{bmatrix}$$
has full column rank.
5. [BK11] If $V_\ell(X, S)$ has rank $\tilde{k} < k$, then there is a minimal pair $(\tilde{X}, \tilde{S}) \in \mathbb{C}^{n\times\tilde{k}} \times \mathbb{C}^{\tilde{k}\times\tilde{k}}$ such that $\mathrm{Span}(\tilde{X}) = \mathrm{Span}(X)$ and $\mathrm{Span}(V_\ell(\tilde{X}, \tilde{S})) = \mathrm{Span}(V_\ell(X, S))$.
6. If (X, S) is a minimal invariant pair, then $(XZ, Z^{-1}SZ)$ is also a minimal invariant pair for every invertible matrix $Z \in \mathbb{C}^{k\times k}$.
7. Let (X, S) be a minimal invariant pair, and let $p_j \in \Pi_k$ be the Hermite interpolating polynomials of f_j at the spectrum of S of maximum degree k. Then (X, S) is a minimal invariant pair of $P(\lambda)\mathbf{x} := \sum_{j=1}^m p_j(\lambda)A_j\mathbf{x} = 0$.
8. Let $(\lambda_j, \mathbf{x}_j)$, $j = 1, \ldots, k$ be eigenpairs of $T(\cdot)$ with $\lambda_i \neq \lambda_j$ for $i \neq j$. Then the invariant pair $(X, S) := ([\mathbf{x}_1, \ldots, \mathbf{x}_k], \mathrm{diag}(\lambda_1, \ldots, \lambda_k))$ is minimal.
9. Consider the nonlinear matrix operator
$$\mathbb{T} \; : \; \begin{cases} \mathbb{C}^{n\times k} \times \mathbb{C}_\Omega^{k\times k} \to \mathbb{C}^{n\times k} \\ (X, S) \mapsto \sum_{j=1}^m A_j X f_j(S) \end{cases} \tag{115.8}$$
where $\mathbb{C}_\Omega^{k\times k}$ denotes the set of $k \times k$ matrices with eigenvalues in Ω. Then an invariant pair (X, S) satisfies $\mathbb{T}(X, S) = 0$, but this relation is not sufficient to characterize (X, S).
To define a scaling condition, choose ℓ such that the matrix $V_\ell(X, S)$ has rank k, and define the partition
$$W = \begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_{\ell-1} \end{bmatrix} := V_\ell(X, S) \left(V_\ell(X, S)^* V_\ell(X, S)\right)^{-1} \in \mathbb{C}^{nk\times k}$$

with $W_j \in \mathbb{C}^{n\times k}$. Then $\mathbb{V}(X,S) = 0$ for the operator

$$\mathbb{V} : \mathbb{C}^{n\times k} \times \mathbb{C}_{\Omega}^{k\times k} \to \mathbb{C}^{n\times k},\ \mathbb{V}(X,S) := W^* V_\ell(X,S) - I_k.$$

If (X,S) is a minimal invariant pair for the nonlinear eigenvalue problem $T(\cdot)\mathbf{x} = 0$, then (X,S) is simple if and only if the linear matrix operator

$$\mathbb{L} : \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k} \to \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k}, (\Delta X, \Delta S) \mapsto (\mathbb{D}\mathbb{T}(\Delta X, \Delta S), \mathbb{D}\mathbb{V}(\Delta X, \Delta S))$$

is invertible, where $\mathbb{D}\mathbb{T}$ and $\mathbb{D}\mathbb{V}$ denotes the Fréchet derivative of $\mathbb{T}$ and $\mathbb{V}$, respectively.

10. [Kre09] The last Fact motivates to apply Newton's method to the system $\mathbb{T}(X,S) = 0$, $\mathbb{V}(X,S) = 0$ which can be written as

$$(X_{p+1}, S_{p+1}) = (X_p, S_p) - \mathbb{L}^{-1}(\mathbb{T}(X_p, S_p), \mathbb{V}(X_p, S_p))$$

where $\mathbb{L} = (\mathbb{D}\mathbb{T}, \mathbb{D}\mathbb{V})$ is the Jacobian matrix of $\mathbb{T}(X,S) = 0, \mathbb{V}(X,S) = 0$.

$$\mathbb{D}\mathbb{T}(\Delta X, \Delta S) = \mathbb{T}(\Delta X, S) + \sum_{j=1}^{m} A_j X[\mathbb{D}f_j(S)](\Delta S),$$

$$\mathbb{D}\mathbb{V}(\Delta X, \Delta S) = W_0^* \Delta X + \sum_{j=1}^{m} W_j^*(\Delta X S^j + X[\mathbb{D}S^j](\Delta S)).$$

Algorithm 8: Newton's method for computing invariant pairs

Require: Initial pair $(X_0, S_0) \in \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k}$ such that $V_\ell(X_0,S_0)^* V_\ell(X_0,S_0) = I_k$

1: $p \leftarrow 0$, $W \leftarrow V_\ell(X_0, S_0)$
2: **repeat**
3: Res $\leftarrow \mathbb{T}(X_p, S_p)$
4: Solve linear matrix equation $\mathbb{L}(\Delta X, \Delta S) = (\text{Res}, O)$
5: $\tilde{X}_{p+1} \leftarrow X_p - \Delta X$, $\tilde{S}_{p+1} \leftarrow S_p - \Delta S$
6: Compute compact QR decomposition $V_\ell(\tilde{X}_{p+1}, \tilde{S}_{p+1}) = WR$
7: $X_{p+1} \leftarrow \tilde{X}_{p+1} R^{-1}$, $S_{p+1} \leftarrow R\tilde{S}_{p+1} R^{-1}$
8: **until** convergence

11. [Bey12, BEK11]

$$\mathbb{T}(X,S) = \int_{\Gamma} T(z) X (zI - S)^{-1}\, dz$$

where Γ is a contour (i.e. a simply closed curve) in Ω containing the spectrum of S in its interior.

12. [BEK11]

$$\mathbb{D}\mathbb{T}(X,S)(\Delta X, \Delta S) = \frac{1}{2\pi i} \int_{\Gamma} (\Delta X + X(zI - S)^{-1}\Delta S)(zI - S)^{-1}\, dz.$$

13. [Eff12] Let (X,S) be a minimal (index ℓ) invariant pair of $T(\cdot)$. If $(\left[\begin{smallmatrix} Y \\ V \end{smallmatrix}\right], M)$ is a minimal invariant pair of the augmented analytic matrix function

$$\hat{T} : \begin{cases} \Omega \to \mathbb{C}^{(n+k)\times(n+k)}, \\ \hat{T}(\mu) \begin{bmatrix} y \\ v \end{bmatrix} = \begin{bmatrix} \mathbb{T}([X,y], \left[\begin{smallmatrix} S & v \\ 0 & \mu \end{smallmatrix}\right]) \\ [V_{\ell+1}^p(X,S)]^* V_{\ell+1}^p([X,y], \left[\begin{smallmatrix} S & v \\ 0 & \mu \end{smallmatrix}\right]) \end{bmatrix} e_{k+1} \end{cases}$$

with $\mathbb{T}$ as in Fact 11, $V_{\ell+1}^p$ analogous to Fact 4, and $e_{k+1} = (0, \ldots, 0, 1)^T \in \mathbb{R}^{k+1}$, then $([X, Y], \left[\begin{smallmatrix} S & V \\ 0 & M \end{smallmatrix}\right])$ is a minimal invariant pair of $T(\cdot)$. Conversely, for any minimal invariant pair $([X, Y], \left[\begin{smallmatrix} S & V \\ 0 & M \end{smallmatrix}\right])$ of $T(\cdot)$ there exists a unique F such that $(\left[\begin{smallmatrix} Y-XF \\ V-(SF-FM) \end{smallmatrix}\right], M)$ is a minimal invariant pair of $\hat{T}(\cdot)$.

14. The previous fact suggests that working with $\hat{T}(\cdot)$ deflates the minimal invariant pair (X, S) from $T(\cdot)$.
15. [Eff12] Effenberger combined the deflation in Fact 13 with the Jacobi–Davidson method to determine several eigenpairs of a nonlinear eigenvalue problem one after another in a safe way.
16. [GKS93] The pair (X, S) is minimal if and only if

$$\operatorname{rank} \begin{bmatrix} \lambda I - S \\ X \end{bmatrix}$$

has full rank for every $\lambda \in \mathbb{C}$ (or, equivalently, for every eigenvalue λ of S).

17. [GKS93] Let $\hat{\lambda}$ be an eigenvalue of $T(\cdot)$ and $X := [\mathbf{x}_0, \ldots, \mathbf{x}_{k-1}] \in \mathbb{C}^{k\times k}$ with $\mathbf{x}_0 \neq 0$. Then $\mathbf{x}_0, \ldots, \mathbf{x}_{k-1}$ is a Jordan chain at $\hat{\lambda}$ if and only if $(X, J_k(\hat{\lambda}))$ is an invariant pair of $T(\cdot)$, where $J_k(\hat{\lambda})$ denotes a $k \times k$ Jordan block corresponding to $\hat{\lambda}$.
18. [BEK11] Let $\hat{\lambda}$ be an eigenvalue of $T(\cdot)$ and consider a matrix $X = [X^{(1)}, \ldots, X^{(p)}]$, $X^{(i)} = [\mathbf{x}_0^{(i)}, \ldots, \mathbf{x}_{m_i}^{(i)}]$, with $\mathbf{x}_0^{(i)} \neq 0$. Then every $\mathbf{x}_0^{(i)}, \ldots, \mathbf{x}_{m_i}^{(i)}$ for $i = 1, \ldots, p$ is a Jordan chain if and only if (X, J) with $J := \operatorname{diag}(J_{m_1}(\hat{\lambda}), \ldots, J_{m_p}(\hat{\lambda}))$ is an invariant pair of $T(\cdot)$. Moreover, (X, J) is minimal if and only if $\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_0^{(p)}$ are linearly independent.
19. [SX12] Suppose that (X, S) is a simple invariant pair of (115.7), $\hat{\lambda}$ an eigenvalue of S, and $J = Z^{-1}SZ$ is the Jordan canonical form of S. Assume that J has m Jordan blocks corresponding to $\hat{\lambda}$, each of size $k_i \times k_i$, $1 \leq i \leq m$. Then there are exactly m Jordan chains of $T(\cdot)$ corresponding to $\hat{\lambda}$, the length of each is k_i, and the geometric multiplicity of $\hat{\lambda}$ is m.
    This fact demonstrates that the spectral structure of an eigenvalue $\hat{\lambda}$ of a matrix function $T(\cdot)$, including the algebraic, partial and geometric multiplicities together with all Jordan chains, is completely represented in a simple invariant pair (X, S) for which $\hat{\lambda}$ is an eigenvalue of S.

Examples:

1. For the quadratic eigenvalue problem (115.2) with eigenvalue $\hat{\lambda} = -1$ and eigenvector $\mathbf{x} = [1; 1]$ the pair $(X, S) := (\mathbf{x}, \hat{\lambda})$ is a minimal invariant pair with minimality index 1, which is not simple, because the algebraic multiplicity of $\hat{\lambda}$ is 2 as an eigenvalue of $T(\lambda)x = 0$ and only 1 as an eigenvalue of S.
    The Jordan pair $(X1, S1)$ with $X1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $S1 = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$ is a minimal invariant pair with minimality index 2, which is simple, and the same is true for the pairs $(X2, S2)$ with $X2 = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$ and $S1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, and $(X3, S3)$ with $X3 := [X1, X2]$ and $S3 := \operatorname{diag}(X1, X2)$.

115.8 The infinite Arnoldi method

Let $T : \Omega \to \mathbb{C}^{n\times n}$ be analytic on a neighborhood Ω of the origin, and assume that $\lambda = 0$ is not an eigenvalue of $T(\cdot)$. To determine eigenvalues close to 0 [JMM12] use the equivalence

of $T(\lambda)\mathbf{x} = 0$ to a linear, infinite dimensional eigenvalue problem, and apply the linear Arnoldi method, which can be reformulated to an iteration involving only standard linear algebra operations on matrices and vectors of finite dimension.

Definitions:

$B(\lambda) := T(0)^{-1}(T(0) - T(\lambda))/\lambda$ for $\lambda \neq 0$ and $B(0) := -T(0)^{-1}T'(0)$ is analytic on Ω, and $\hat{\lambda}$ is an eigenvalue of $T(\cdot)$ if and only if $\hat{\lambda}$ is an eigenvalue of $\lambda B(\lambda)\mathbf{x} = \mathbf{x}$.

Let $D(\mathcal{B}) := \{\phi \in C^\infty(\mathbb{R}, \mathbb{C}^n) : \sum_{i=0}^\infty B^{(i)}(0)\phi^{(i)}(0) < \infty\}$, and define

$$\mathcal{B}(\theta) := \int_0^\theta \phi(\hat{\theta})\, d\hat{\theta} + C(\phi), \quad C(\phi) := \sum_{i=0}^\infty \frac{B^{(i)}(0)}{i!}\phi^{(i)}(0) = \left(B(\frac{d}{d\theta})\phi\right)(0). \tag{115.9}$$

$(\Psi, R) \in D(\mathcal{B})^p \times \mathbb{C}^{p\times p}$ is an **invariant pair of the operator** $\mathcal{B}$ if $(\mathcal{B}\Psi)(\theta) = \Psi(\theta)R$.

Facts:

The following facts can be found in [JMM12].

1. Let $\mathbf{x} \in \mathbb{C}^n \setminus \{0\}$, $\lambda \in \Omega$ and denote $\phi(\theta) := \mathbf{x}e^{\lambda\theta}$. Then the following two statements are equivalent:
   (i) $(\hat{\lambda}, \hat{\mathbf{x}})$ is an eigenpair of $T(\cdot)$
   (ii) $(\hat{\lambda}, \phi)$ is an eigenpair of the linear, infinite dimensional eigenvalue problem $\lambda\mathcal{B}\phi = \phi$.
2. All eigenfunctions of $\mathcal{B}$ depend exponentially on θ, i.e. if $\lambda\mathcal{B}\psi = \psi$, then $\psi(\theta) = \mathbf{x}e^{\lambda\theta}$.
3. The (linear) Arnoldi method for the operator $\mathcal{B}$ is given in Algorithm 9. Here $\langle\cdot,\cdot\rangle$ denotes a scalar product on $C^\infty(\mathbb{R}, \mathbb{C}^n)$, and $H_k = (h_{ik}) \in \mathbb{C}^{k\times k}$ a Hessenberg matrix constructed in the algorithm.

Algorithm 9: Arnoldi method for $\mathcal{B}$

Require: Initial function ϕ_1 with $\langle\phi_1, \phi_1\rangle = 1$
for $k = 1, 2, \ldots$ until convergence **do**
  $\psi \leftarrow \mathcal{B}\psi_k$
  for i=1,...,k **do**
    $h_{ik} \leftarrow \langle\psi, \phi_i\rangle$
    $\psi \leftarrow \psi - h_{ik}\phi_i$
  end for
  $h_{k+1,k} \leftarrow \sqrt{\langle\psi, \psi\rangle}$
  $\phi_{k+1} \leftarrow \psi/h_{k+1,k}$
end for
Compute eigenvalues μ_i of Hessenberg matrix H_k
Return eigenvalue approximations $1/\mu_i$ of $\mathcal{B}$

4. Since the Arnoldi method favors extreme eigenvalues of $\mathcal{B}$, $1/\mu_i$ will approximate eigenvalues of $T(\cdot)$ close to the origin.
5. If ϕ_1 is a polynomial of degree k, then $\mathcal{B}\phi$ is a polynomial of degree $k+1$. Hence, if ϕ_1 is a constant function, then Algorithm 9 after N steps arrives at a Krylov space $\mathcal{K}_N(\mathcal{B}, \phi_1) = \mathrm{Span}\{\phi_1, \ldots, \phi_N\}$ of vectors of polynomials of degree $N-1$.
6. Let $\{q_i\}_{i=0,1,\ldots}$ be a sequence of polynomials such that q_i is of degree i with non-zero leading coefficient, and let $q_0 \equiv 1$. Let $L_N \in \mathbb{R}^{N\times N}$ be an integration map

corresponding to $\{q_i\}$ such that

$$\begin{bmatrix} q_0(\theta) \\ \vdots \\ q_{N-1}(\theta) \end{bmatrix} = L_N \begin{bmatrix} q_1'(\theta) \\ \vdots \\ q_N'(\theta) \end{bmatrix}.$$

Let the columns of $(\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}) =: X \in \mathbb{C}^{n\times N}$ denote the vector coefficients in the basis $\{q_i\}$, and denote a vector of polynomials $\phi(\theta) := \sum_{i=0}^{N-1} q_i(\theta)\mathbf{x}_i$.
If $\psi(\theta) = (\mathcal{B}\phi)(\theta) =: \sum_{i=0}^{N} q_i(\theta)\mathbf{y}_i$, then the coefficients $\mathbf{y}_i$ of $\mathcal{B}\phi$ are given by $(\mathbf{y}_1, \ldots, \mathbf{y}_N) = XL_N$ and $\mathbf{y}_0 = \left(\sum_{i=0}^{N-1} B(\frac{d}{d\theta})q_i(\theta)\mathbf{x}_i\right)(0) - \sum_{i=1}^{N} q_i(0)\mathbf{y}_i$.
This fact permits to reformulate Algorithm 9 to an iteration involving only standard linear algebra operations on matrices and vectors of finite dimension.
In [JMM12] the details are worked out for two polynomial bases, the monomial basis $q_i = \theta^i$ and Chebyshev polynomials.

7. [JMM11] Suppose that $S \in \mathbb{C}^{p\times p}$ is invertible and suppose that (Ψ, S^{-1}) is an invariant pair of $\mathcal{B}$. Then Ψ can be expressed as $\Psi(\theta) = X\exp(\theta S)$ for some matrix $X \in \mathbb{C}^{p\times p}$.
8. [JMM11] Assume that $T(\lambda) := \sum_{j=1}^{m} f_j(\lambda)A_j$. Let $S \in \mathbb{C}^{p\times p}$ be nonsingular and $X \in \mathbb{C}^{n\times p}$. The following two statements are equivalent:
   (i) (Ψ, S^{-1}), $\Psi(\theta) := X\exp(\theta S)$ is an invariant pair of the operator $\mathcal{B}$.
   (ii) (X, S) is an invariant pair of $T(\cdot)$, i.e. $\sum_{j=1}^{m} A_j X f_j(S) = 0$.
9. Inspired by the implicitly restarted Arnoldi method for linear eigenproblems [JMM11] proposes a variant of the infinite Arnoldi method for the nonlinear eigenvalue problem $T(\lambda)x = 0$ which allows for locking already converged eigenpairs. The locked part of the partial Schur factorization for linear problems is replaced by invariant pairs. The method uses functions $\phi(\theta) = Xe^{S\theta}c + q(\theta)$ where $X \in \mathbb{C}^{n\times p}$, $S \in \mathbb{C}^{p\times p}$, $c \in \mathbb{C}^p$ and $q : \mathbb{C} \to \mathbb{C}^n$ is a vector of polynomials.

References

[AR68] P.M. Anselone, L.B. Rall, The solution of characteristic value-vector problems by Newton's method, *Numer. Math.*, 11:38–45, 1968.

[AST09] J. Asakura, T. Sakurai, H. Tadano, T. Ikegami, K. Kimura, A numerical method for nonlinear eigenvalue problems using contour integrals, *JSIAM Letters*, 1:52–55, 2009.

[BV04] T. Betcke, H. Voss, A Jacobi–Davidson–type projection method for nonlinear eigenvalue problems, *Future Generation Comput. Syst.*, 20:363–372, 2004.

[Bey12] W.-J. Beyn, An integral method for solving nonlinear eigenvalue problems, *Linear Algebra Appl.*, 436:3839–3863, 2012.

[BEK11] W.-J. Beyn, C. Effenberger, D. Kressner, Continuation of eigenvalues and invariant pairs for parametrized nonlinear eigenvalue problems, *Numer. Math.*, 119:489–516, 2011.

[BT09] W.-J. Beyn, V. Thümmler, Continuation of low-dimensional invariant subspaces in dynamical systems of large dimension, *SIAM J. Matrix Anal. Appl.*, 31:1361–1381, 2009.

[BK11] T. Betcke, D. Kressner, Perturbation, extraction and refinement of invariant pairs, *Linear Algebra Appl.*, 435:514–536, 2011.

[DP01] E.M. Daya, M. Potier–Ferry, A numerical method for nonlinear eigenvalue problems application to vibrations of viscoelastic structures, *Comupers & Structures*, 79:533–541, 2001.

[Duf55] R.J. Duffin, A minimax theory for overdamped networks, *J. Rat. Mech. Anal.*, 4:221–233, 1955.

[Eff12] C. Effenberger, Robust successive computation of eigenpairs for nonlinear eigenvalue problems, Tech.Report 27.2012, Math. Inst. of Comput. Sc. Engn., EPF Lausanne 2012.

[GKS93] I. Gohberg, M.A. Kaashoek, F. van Schagen, On the local theory of regular analytic functions, *Linear Algebra Appl.*, 182:9–25, 1993.

[GLR82] I. Gohberg, P. Lancaster, L. Rodman. *Matrix Polynomials*, Academic Press, New York, 1982.

[GR81] I. Gohberg, L. Rodman, On the local theory of regular analytic functions, *Linear Algebra Appl.*, 182:9–25, 1993.

[Had68] K.P. Hadeler, Variationsprinzipien bei nichtlinearen Eigenwertaufgaben, *Arch. Ration. Mech. Anal.*, 30:297–307, 1968.

[HS03] M.E. Hochstenbach, G.L.P Sleijpen, Two-sided and alternating Jacobi-Davidson,, *Linear Algebra Appl.*,358:145–172, 2003.

[HL99] V. Hryniv, P. Lancaster, On the perturbation of analytic matrix functions, *Integral Equations Operator Theory*,34:325–338, 1999.

[Jar12] E. Jarlebring, Convergence factor for Newton methods for nonlinear Eigenvalue problems, *Linear Algebra Appl.*,436:3943–3853, 2012.

[JMM11] E. Jarlebring, W. Michiels, K. Meerbergen, Computing a partial Schur factorization of nonlinear eigenvalue problems using the infinite Arnoldi method, TechRep Dept. Comp. Science, K.U. Leuven, 2011.

[JMM12] E. Jarlebring, W. Michiels, K. Meerbergen, A linear eigenvalue algorithm for the nonlinear eigenvalue problem, *Numer. Math*, accepted, 2012.

[Kre09] D. Kressner, A block Newton method for nonlinear eigenvalue problems, *Numer. Math.*,114:355–372, 2009.

[Kub70] 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.

[Lan61] P. Lancaster, A generalised Rayleigh quotient iteration for lambda-matrices, *Arch. Rat. Mech. Anal.*,8:309–322, 1961.

[LBL10] B.-S. Liao, Z. Bai, L.-Q. Lee, K. Ko, Nonlinear Rayleigh–Ritz iterative method for solving large scale nonlinear eigenvalue problems, *Taiwanese J. Math.*, 14:869–883, 2010.

[Mee01] K. Meerbergen, Locking and restarting quadratic eigenvalue solvers, *SIAM J. Sci. Comput.*, 22:1814–1839, 2001.

[Neu85] A. Neumaier, Residual inverse iteration for the nonlinear eigenvalue problem, *SIAM J. Numer. Anal.*,22:914–923, 1985.

[NV10] V. Niendorf, H. Voss, Detecting hyperbolic and definite matrix polynomials, *Linear Algebra Appl.*,432:1017–1035, 2010.

[Rog64] E.H. Rogers, A minimax theory for overdamped systems, *Arch. Ration. Mech. Anal.*,16:89–96, 1964.

[Rot89] K. Rothe. *Lösungsverfahren für nichtlineare Matrixeigenwertaufgaben mit Anwendungen auf die Ausgleichselementmethode*, Ph.D. Thesis, Universität Hamburg, Germany, 1989.

[Ruh73] A. Ruhe, Algorithms for nonlinear eigenvalue problems, *SIAM J. Numer. Anal.*,10:674–689, 1973.

[Sch08] K. Schreiber. *Nonlinear Eigenvalue Problems: Newton-type Methods and Nonlinear Rayleigh Functionals*, Ph.D. Thesis, TU Berlin, Germany, 2008.

[SS06] H. Schwetlick, K. Schreiber, A primal-dual Jacobi–Davidson-like method for nonlinear eigenvalue problems, TechRep ZIH-IR-0613, Technische Universität Dresden, Germany, 2006.

[SS10] H. Schwetlick, K. Schreiber, Nonlinear Rayleigh functionals. *Linear Algebra Appl.*, 436:3991 – 4016, 2012.

[SBF96] G.L. Sleijpen, G.L. Booten, D.R. Fokkema, H.A. van der Vorst, Jacobi-Davidson type methods for generalized eigenproblems and polynomial eigenproblems, *BIT Numerical*

Mathematics, 36:595–633, 1996.

[SX11] D. Szyld, F. Xue, Local convergence analysis of several inexact Newton-type algorithms for general nonlinear eigenvalue problems, TechRep 11-08-09, Temple University, Philadelphia , USA, 2011.

[SX12] D. Szyld, F. Xue, Several properties of invariant pairs of nonlinear algebraic eigenvalue problems TechRep 12-02-09, Temple University, Philadelphia , USA, 2012.

[Vos03] H. Voss, A maxmin principle for nonlinear eigenvalue problems with application to a rational spectral problem in fluid–solid vibration, *Appl. Math.*, 48:607–622, 2003.

[Vos04] H. Voss, An Arnoldi method for nonlinear eigenvalue problems, *BIT Numerical Mathematics*, 44:387–401, 2004.

[Vos07] H. Voss, A new justification of the Jacobi–Davidson method for large eigenproblems, *Linear Algebra Appl.*, 424:448–455, 2007.

[Vos07a] H. Voss, A Jacobi–Davidson method for nonlinear and nonsymmetric eigenproblems, *Computers & Structures*, 85:1284–1292, 2007.

[Vos09] H. Voss, A minmax principle for nonlinear eigenproblems depending continuously on the eigenparameter, *Numer. Linear Algebra Appl.*, 16:899–913, 2009.

[VW82] H. Voss, B. Werner, A minimax principle for nonlinear eigenvalue problems with applications to nonoverdamped systems, *Math. Meth. Appl. Sci.*, 4:415–424, 1982.

[Wer70] B. Werner. *Das Spektrum von Operatorenscharen mit verallgemeinerten Rayleighquotienten*, Ph.D. Thesis, Universität Hamburg, Germany, 1970.

[YMW07] C. Yang, J.C. Meza, L.-W. Wang, A trust region direct constrained minimization algorithm for the Kohn-Sham equation, *SIAM J. Sci. Comput.*, 29:1854–1875, 2007.