Perturbation, Extraction and Refinement of Invariant Pairs for Matrix Polynomials

Timo Betcke* Daniel Kressner†

February 2, 2010

Abstract

Generalizing the notion of an eigenvector, invariant subspaces are frequently used in the context of *linear* eigenvalue problems, leading to conceptually elegant and numerically stable formulations in applications that require the computation of several eigenvalues and/or eigenvectors. Similar benefits can be expected for *polynomial* eigenvalue problems, for which the concept of an invariant subspace needs to be replaced by the concept of an invariant pair. Little has been known so far about numerical aspects of such invariant pairs. The aim of this paper is to fill this gap. The behavior of invariant pairs under perturbations of the matrix polynomial is studied and a first-order perturbation expansion is given. From a computational point of view, we investigate how to best extract invariant pairs from a linearization of the matrix polynomial. Moreover, we describe efficient refinement procedures directly based on the polynomial formulation. Numerical experiments with matrix polynomials from a number of applications demonstrate the effectiveness of our extraction and refinement procedures.

1 Introduction

Given a matrix polynomial

$$P(\lambda) = A_0 + \lambda A_1 + \lambda^2 A_2 + \cdots + \lambda^\ell A_\ell \tag{1}$$

with $n \times n$ matrices $A_0, \ldots, A_\ell$, a vector $x \neq 0$ is called an *eigenvector* belonging to some *eigenvalue* λ_0 of P if $P(\lambda_0)x = 0$. Generalizing the notion of an *eigenpair* (x, λ), a pair $(X, S) \in \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k}$ is called *invariant* if the relation

$$\mathbb{P}(X, S) := A_0X + A_1XS + A_2XS^2 + \cdots + A_\ell XS^\ell = 0. \tag{2}$$

is satisfied. One could regard the space $\mathcal{X}$ spanned by the columns of X as an invariant subspace for P. However, as we will see in the course of this paper, the notion of invariant subspaces is rather inconvenient when dealing with polynomial eigenvalue problems and the notion of invariant pairs should be preferred.

*t.betcke@reading.ac.uk, Department of Mathematics, University of Reading, UK. Timo Betcke acknowledges support by Engineering and Physical Sciences Research Council grant EP/D079403/1.

†kressner@math.ethz.ch, Seminar for applied mathematics, ETH Zurich, Switzerland.

For linear eigenvalue problems, it is well known that working with invariant subspaces instead of eigenvectors offers conceptual and numerical benefits [16]. For example, eigenvectors associated with a multiple eigenvalue are unstable under perturbations, that is, an arbitrarily small change in the matrix may cause some of the eigenvectors disappear. In contrast, the corresponding invariant subspace remains stable under perturbations, provided that it is simple, that is, the algebraic eigenvalue multiplicities of the invariant subspace coincide with those of the matrix. It will be seen that similar statements hold for matrix polynomials; working with invariant pairs generally increases the robustness of numerical methods in the presence of (nearly) multiple eigenvalues.

For $k = n\ell$, invariant pairs are closely connected to the notion of standard pairs developed by Gohberg, Lancaster, and Rodman [15]. For $k < n\ell$, invariant pairs could therefore be seen as local versions of standard pairs. If S is in Jordan canonical form then (X, S) is called a Jordan pair. As the focus of this paper is on numerical aspects, we shall not discuss this connection in more detail.

For $k = n$ and invertible X, any matrix S satisfying (2) gives rise to a solvent XSX^{-1} for the polynomial P defined in (1). We refer to Higham and Kim [19] for existing results on solvents for $\ell = 2$. Currently, it is not clear to us how solvents can be put to good use in the context of invariant pairs. One emphasis of this paper is that it is best, both from a theoretical and numerical point of view, to consider the matrices X and S (or XSX^{-1}) not as separate entitities but only jointly in an invariant pair (X, S).

For $k = 1$, invariant pairs coincide with eigenpairs (provided that $X \neq 0$). Numerical aspects of eigenpairs for matrix polynomials have been studied quite intensively in the last decade. A number of theoretical results concerning the sensitivity of eigenvalues and eigenvectors of matrix polynomials under (structured) perturbations are available [5, 11, 1].

The polynomial eigenvalue problem (1) is usually solved via linearization and a large class of linearizations particularly suitable for computing eigenpairs has been introduced in Mackey et al. [29]. The effects of linearization on the (structured) eigenvalue sensitivity and backward error have been studied in [20, 21, 1], leading to clear recommendations which linearization is to preferred from a numerical point of view. Scaling and balancing are preprocessing steps that aim at improving the accuracy of computed eigenpairs, see [6, 13, 22].

The purpose of this paper is to discuss numerical aspects of invariant pairs for general k. Little is known in this direction so far, with the notable exception of the work by Beyn and Thümmler [9] on the continuation of invariant pairs for monic quadratic matrix polynomials. In fact, the work on this paper was very much inspired by the results in [9] and we will point out connections whenever possible.

The rest of this paper is organized as follows. Section 2 is concerned with basic properties of invariant pairs and introduces the notions of minimality and simplicity. In Section 3, we study the first-order behavior of an invariant pair under perturbations of the matrix polynomial. In particular, Theorem 7 reveals that simple invariant pairs combined with a suitable normalization condition are well-posed. Section 4 investigates computational aspects and presents several approaches to extracting invariant pairs from the solution of the linearized eigenvalue problem. Numerical experiments suggest that a novel approach based on the generalized singular value decomposition is the preferred one. In Section 5, we describe a Newton iteration for refining invariant pairs and investigate the solution of the corresponding linearized equations in some detail. Section 6 contains some numerical experiments demonstrating the use of the presented concepts and algorithms in applications. Appendix A serves to illustrate the relation between Jordan chains for matrix polynomials and invariant pairs.

Remark 1 *Recent numerically oriented work on polynomial eigenvalue problems, see for example [20, 21], has shifted towards the use of a homogeneous formulation* $P(\alpha,\beta) = \beta^\ell A_0 + \alpha\beta^{\ell-1}A_1 + \alpha^2\beta^{\ell-2}A_2 + \cdots + \alpha^\ell A_\ell$ *in place of (1), partly because it elegantly allows for the simultaneous treatment of finite and infinite eigenvalues. At least for* $\ell = 1$*, it is known how to put invariant subspaces in a homogeneous framework: by using pairs of deflating subspaces [35, 36]. However, it is not clear how to extend the concept of deflating subspaces to matrix polynomials. The notion of decomposable pairs from Chapter 7 in [15] does not appear to be suitable for this purpose as decomposability still relies on a strict separation between finite and infinite eigenvalues.*

It should be emphasized, however, that infinite eigenvalues can still be covered by defining invariant pairs for the reverse polynomial, similar to the concept of infinite Jordan pairs from [15]. The only restriction imposed by such an approach is that an invariant pair may not contain both, zero and infinite eigenvalues simultaneously. If a polynomial has zero and infinite eigenvalues, they have to be handled by separate invariant pairs, one for the original and one for reverse polynomial.

2 Preliminaries

In this section, we provide basic theoretical results on invariant pairs for matrix polynomials. Throughout this paper, we only consider matrix polynomials that are regular: $\det(P(\lambda)) \not\equiv 0$.

The definition of an invariant pair (2) is independent of the choice of basis. To see this, let $T \in \mathbb{C}^{k\times k}$ be an invertible matrix and consider $\tilde{X} = XT$. Then multiplying (2) with T from the right yields

$$A_0\tilde{X} + A_1\tilde{X}\tilde{S} + A_2\tilde{X}\tilde{S}^2 + \cdots + A_\ell\tilde{X}\tilde{S}^\ell = 0, \quad \tilde{S} = T^{-1}ST, \tag{3}$$

and hence $(\tilde{X}, \tilde{S})$ is also an invariant pair. If S is diagonalizable then T can be chosen such that

$$\tilde{S} = T^{-1}ST = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k).$$

In this case the relation (3) implies that the columns $\tilde{x}_1, \ldots, \tilde{x}_k$ of the transformed basis $\tilde{X}$ are eigenvectors of P: $P(\lambda_i)\tilde{x}_i = 0$, provided of course that $\tilde{x}_i \neq 0$. This shows that the eigenvalues of S form a subset of the eigenvalues of P. More generally, if $\tilde{S}$ is in Jordan canonical form then the columns of $\tilde{X}$ contain Jordan chains for P [15, Proposition 1.10], see also Appendix A.

2.1 Simple invariant pairs and deflating subspaces

In contrast to linear eigenvalue problems, eigenvectors belonging to mutually distinct eigenvalues are not necessarily linearly independent. For example, the matrix polynomial [12]

$$P(\lambda) = \begin{bmatrix} 0 & 12 \\ -2 & 14 \end{bmatrix} + \lambda \begin{bmatrix} -1 & -6 \\ 2 & -9 \end{bmatrix} + \lambda^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

has the same eigenvector $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ belonging to the eigenvalues 3 and 4. Hence, a given full rank matrix X that is known to be part of an invariant pair may not uniquely determine the matrix S such that (X, S) is an invariant pair. It is not even reasonable to require X to have full rank. These limitations raise doubts whether the concept of an invariant subspace (i.e., the

space spanned by the columns of X) is appropriate at all for polynomial eigenvalue problems and we therefore favor the concept of an invariant pair.

To allow for rank deficiencies in X, the following notion of minimality will be used, which has first been proposed in [9] for $\ell = 2$.

Definition 2 (Minimal pair) *A pair $(X, S) \in \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k}$ is called* minimal *if there is $m \in \mathbb{N}$ such that*

$$V_m(X, S) := \begin{bmatrix} XS^{m-1} \\ \vdots \\ XS \\ X \end{bmatrix} \tag{4}$$

has full column rank. The smallest such m is called minimality index of (X, S).

By the Cayley-Hamilton theorem, the minimality index of a minimal pair cannot exceed k, see also [28, Lemma 5]. Moreover, it will be shown in Lemma 5 below that the minimality index cannot exceed the degree of the matrix polynomial.

The following theorem shows that it is always possible to extract a minimal invariant pair with minimality index at most ℓ from a non-minimal one. This allows us to restrict the discussion in this paper to minimal invariant pairs.

Theorem 3 *Let (X, S) be an invariant pair for a matrix polynomial P of degree ℓ. Then there is a minimal invariant pair $(\tilde{X}, \tilde{S})$ with minimality index at most ℓ such that*

$$\text{span } V_\ell(\tilde{X}, \tilde{S}) = \text{span } V_\ell(X, S),$$

with $V_\ell(X, S)$ and $V_\ell(\tilde{X}, \tilde{S})$ defined as in (4).

Proof. Let $\tilde{k}$ denote the rank of $V_\ell(X, S)$. If (X, S) is not minimal, $\tilde{k} < k$ and after a change of basis we may assume that the null space of $V_\ell(X, S)$ is spanned by the unit vectors $e_{\tilde{k}+1}, \ldots, e_k$. This implies that the last $k - \tilde{k}$ columns of $X, XS, \ldots, XS^{\ell-1}$ are zero. Let us partition

$$X = \begin{bmatrix} \tilde{X}, 0 \end{bmatrix}, \quad S = \begin{bmatrix} \tilde{S} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

with $\tilde{X} \in \mathbb{C}^{n\times\tilde{k}}$ and $\tilde{S} \in \mathbb{C}^{\tilde{k}\times\tilde{k}}$. Then, by induction,

$$\begin{aligned} XS &= \begin{bmatrix} \tilde{X}\tilde{S}, 0 \end{bmatrix} \\ XS^2 &= \begin{bmatrix} \tilde{X}\tilde{S}, 0 \end{bmatrix} S = \begin{bmatrix} \tilde{X}\tilde{S}^2, 0 \end{bmatrix} \\ &\vdots \\ XS^{\ell-1} &= \begin{bmatrix} \tilde{X}\tilde{S}^{\ell-2}, 0 \end{bmatrix} S = \begin{bmatrix} \tilde{X}\tilde{S}^{\ell-1}, 0 \end{bmatrix} \\ XS^{\ell} &= \begin{bmatrix} \tilde{X}\tilde{S}^{\ell-1}, 0 \end{bmatrix} S = \begin{bmatrix} \tilde{X}\tilde{S}^{\ell}, \star \end{bmatrix}. \end{aligned}$$

Hence, the first $\tilde{k}$ columns of the relation $\mathbb{P}(X, S) = 0$ amount to $\mathbb{P}(\tilde{X}, \tilde{S}) = 0$, showing that $(\tilde{X}, \tilde{S})$ is an invariant pair for P. By construction, $V_\ell(\tilde{X}, \tilde{S})$ has full column rank and thus $(\tilde{X}, \tilde{S})$ is minimal. □

An eigenvector x of P is called *simple* if the corresponding eigenvalue λ_0 is a simple root of $\det(P(\lambda))$. The following definition provides an appropriate extension of this concept to invariant pairs, see also [9].

Definition 4 (Simple invariant pair) *An invariant pair (X, S) for a regular matrix polynomial P of degree ℓ is called* simple *if (X, S) is minimal and the algebraic multiplicities of the eigenvalues of S are identical to the algebraic multiplicities of the corresponding eigenvalues of P.*

The definition of invariant pairs is motivated by their connection to standard and generalized eigenvalue problems via the *companion form linearization*

$$C(\lambda) = C_{\mathcal{A}} + \lambda C_{\mathcal{B}} = \begin{bmatrix} A_{\ell-1} & A_{\ell-2} & \cdots & A_0 \\ -I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -I_n & 0 \end{bmatrix} + \lambda \begin{bmatrix} A_\ell & 0 & \cdots & 0 \\ 0 & I_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_n \end{bmatrix}. \tag{5}$$

The eigenvalues of $C_{\mathcal{A}} + \lambda C_{\mathcal{B}}$ are identical with the eigenvalues of P. In particular, the regularity of P implies the regularity of $C_{\mathcal{A}} + \lambda C_{\mathcal{B}}$. Moreover, if (X, S) is an invariant pair and A_ℓ is invertible then it is easy to see that $\operatorname{span}(V_\ell(X, S))$ is an invariant subspace for $C_{\mathcal{B}}^{-1} C_{\mathcal{A}}$. For the more general case, where A_ℓ may be singular, we note that

$$C_{\mathcal{A}}\, V_\ell(X, S) = \begin{bmatrix} \sum_{j=0}^{\ell} A_j X S^j \\ -X S^\ell \\ \vdots \\ -X S^2 \\ -X S \end{bmatrix} = \begin{bmatrix} -A_\ell X S^\ell \\ -X S^{\ell-1} \\ \vdots \\ -X S^2 \\ -X S \end{bmatrix}, \quad C_{\mathcal{B}}\, V_\ell(X, S) = \begin{bmatrix} A_\ell X S^{\ell-1} \\ X S^{\ell-2} \\ \vdots \\ X S \\ X \end{bmatrix}. \tag{6}$$

This shows $C_{\mathcal{A}}\, V_\ell(X, S) + C_{\mathcal{B}}\, V_\ell(X, S)\, S = 0$ and hence $(V_\ell(X, S), S)$ is a minimal invariant pair for the matrix pencil $C_{\mathcal{A}} + \lambda C_{\mathcal{B}}$.[1] Note that Lemma 5 below implies that actually $V_\ell(X, S)$ itself has full rank and therefore its minimality index is 1. Later on, in Section 4, we will see that the opposite direction of the above derivations is also possible; we can always extract invariant pairs for P from simple invariant pairs for $C_{\mathcal{A}} + \lambda C_{\mathcal{B}}$.

Lemma 5 *Let (X, S) be a minimal invariant pair of a regular matrix polynomial of degree ℓ. Then the minimality index of (X, S) does not exceed ℓ.*

Proof. Suppose that the minimality index is larger than ℓ. Then there is $v \neq 0$ such that $Xv = XSv = \cdots = XS^{k-1}v = 0$ and $XS^k v \neq 0$ for some $k \geq \ell$. By the invariance of (X, S),

$$\sum_{j=0}^{\ell} A_j X S^j = 0 \;\Rightarrow\; \sum_{j=0}^{\ell} A_j X S^{j+k-\ell} = 0 \;\Rightarrow\; \sum_{j=0}^{\ell} A_j X S^{j+k-\ell} v = 0,$$

and hence $A_\ell X S^k v = 0$. This implies that the vector

$$y = V_\ell(X, S) S^{1+k-\ell} v = \begin{bmatrix} X S^k v \\ 0 \\ \vdots \\ 0 \end{bmatrix} \neq 0$$

[1] In the usual language of matrix pencils [36], one would call $\operatorname{span}(V_\ell(X, S))$ a right deflating subspace belonging to the eigenvalues of S. To stay notationally consistent we will often use the concept of invariant pairs also in the linear case.

satisfies $C_{\mathcal{B}} y = 0$ and hence y is an eigenvector belonging to the eigenvalue ∞ of the companion matrix pencil $C_{\mathcal{A}} + \lambda C_{\mathcal{B}}$. On the other hand, by its definition y is also contained in the deflating subspace $\operatorname{span}\big(V_\ell(X,S)\big)$ belonging to eigenvalues of S. Hence, the intersection of the deflating subspace belonging to the eigenvalue ∞ and the deflating subspace belonging to the (finite) eigenvalues of S is nontrivial. By standard results for matrix pencils [36] this is not possible since $C_{\mathcal{A}} + \lambda C_{\mathcal{B}}$ is regular according to the assumption. □

Lemma 6 *A minimal invariant pair (X,S) for a regular matrix polynomial is simple if and only if $(V_\ell(X,S),S)$ is a simple invariant pair for the corresponding companion linearization.*

Proof. This follows directly from the one-to-one correspondence between the eigenvalues of $C_{\mathcal{A}} + \lambda C_{\mathcal{B}}$ and P. □

3 First-order Perturbation Theory

Given a matrix polynomial P of the form (1), let us consider the nonlinear matrix operator

$$\begin{array}{rcl} \mathbb{P} : \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k} & \to & \mathbb{C}^{n\times k}, \\ (X,S) & \mapsto & A_0X + A_1XS + \cdots + A_\ell XS^\ell. \end{array} \tag{7}$$

By definition, a simple invariant pair (X,S) satisfies $\mathbb{P}(X,S) = 0$. As this condition is not sufficient to characterize (X,S) we add the condition $W^{\mathsf{H}} V_m(X,S) = I_k$, where $m \le \ell$ is not smaller than the minimality index of (X,S) and the columns of $W = [W_{m-1}^{\mathsf{H}}, \ldots, W_0^{\mathsf{H}}]^{\mathsf{H}}$ form an orthonormal basis of $\operatorname{span}(V_m(X,S))$. Note that W is considered to be fixed throughout this section.

In the following, we study the change of (X,S) under small perturbations of the coefficients of the polynomial:

$$(P + \triangle P)(\lambda) = (A_0 + E_0) + \lambda(A_1 + E_1) + \cdots + \lambda^\ell (A_\ell + E_\ell) \tag{8}$$

for general matrices $E_0, \ldots, E_\ell \in \mathbb{C}^{n\times n}$. In other words, we look for a nearby pair $(\hat{X}, \hat{S})$ that satisfies the equations

$$(\mathbb{P} + \triangle\mathbb{P})(\hat{X}, \hat{S}) = 0, \quad W^{\mathsf{H}} V_m(\hat{X}, \hat{S}) - I = 0, \tag{9}$$

with $\mathbb{P} + \triangle\mathbb{P}$ defined as in (7) but with perturbed coefficients.

Stewart [34, 35] analyzed perturbations of invariant and deflating subspaces associated with linear eigenvalue problems by solving the corresponding quadratic matrix equations (9) with a fixed point iteration. Apart from pioneering the study of perturbed invariant subspaces for non-normal matrices, Stewart's approach has the additional merit of admitting exact bounds, provided that the norm of the perturbation stays below a certain specified threshold. Although an extension of this approach to polynomial eigenvalue problems would be possible by applying the Newton-Kantorovich theorem to (9), we restrict ourselves to first-order perturbation expansions. Perturbation expansions of first and higher order for invariant subspaces of matrices have been pioneered by Sun [37].

3.1 Solvability of the linearized matrix equations

For the linearization of the nonlinear matrix equations (9), we set $\hat{X} = X + \triangle X$, $\hat{S} = S + \triangle S$ and consider $\|E_j\|_F \le \varepsilon, \|\triangle X\|_F \le \varepsilon$, $\|\triangle S\|_F \le \varepsilon$ for some sufficiently small $\varepsilon > 0$. Omitting terms of order $\mathcal{O}(\varepsilon^2)$ as $\varepsilon \to 0$ the linearized equations read as follows:

$$\mathbb{L}_P(\triangle X, \triangle S) = -\triangle\mathbb{P}(X, S), \quad \mathbb{L}_V(\triangle X, \triangle S) = 0, \tag{10}$$

with

$$\mathbb{L}_P : (\triangle X, \triangle S) \;\mapsto\; \mathbb{P}(\triangle X, S) + \sum_{j=1}^{\ell} A_j X\, \mathbb{D}S^j(\triangle S), \tag{11}$$

$$\mathbb{L}_V : (\triangle X, \triangle S) \;\mapsto\; W_0^{\mathsf{H}} \triangle X + \sum_{j=1}^{m-1} W_j^{\mathsf{H}}\big(\triangle X\, S^j + X\, \mathbb{D}S^j(\triangle S)\big), \tag{12}$$

where $\mathbb{D}S^j$ denotes the Fréchet derivative of the map $S \mapsto S^j$:

$$\mathbb{D}S^j : \triangle S \mapsto \sum_{i=0}^{j-1} S^i\, \triangle S\, S^{j-i-1}. \tag{13}$$

For example, for $\ell = m = 2$, the linear matrix operators (11)–(12) amount to

$$\begin{aligned}
\mathbb{L}_P(\triangle X, \triangle S) &= A_0\, \triangle X + A_1\, \triangle X\, S + A_2\, \triangle X\, S^2 + A_1 X\, \triangle S + A_2 X\big(\triangle S\, S + S\, \triangle S\big),\\
\mathbb{L}_V(\triangle X, \triangle S) &= W_0^{\mathsf{H}} \triangle X + W_1^{\mathsf{H}}\big(\triangle X\, S + X\, \triangle S\big).
\end{aligned}$$

Theorem 7 *Let (X, S) be a minimal invariant pair for a regular matrix polynomial P. Then the linear system of matrix equations (10) has a unique solution $(\triangle X, \triangle S)$ if and only if (X, S) is simple.*

Proof. For the case $\ell = 2$ *and* invertible A_ℓ, this result is proven in [9, Thm 2.2] based on results from [8]. The extension of the proof to $\ell \neq 2$ is relatively easy but the extension to singular A_ℓ requires a more significant change.

We first note that $m = \ell$ can be assumed without loss of generality. If $m < \ell$ we simply define $\tilde{W}$ to be W padded with zeros such that $W^{\mathsf{H}} V_m(\hat{X}, \hat{S}) = \tilde{W}^{\mathsf{H}} V_\ell(\hat{X}, \hat{S})$ and work with the latter formulation.

By Lemma 6, (X, S) is simple if and only if $(V_\ell(X, S), S)$ is a simple invariant pair for the companion linearization $C_{\mathcal{A}} + \lambda C_{\mathcal{B}}$ defined in (5). By existing results on generalized eigenvalue problems [27, 36], the latter condition is equivalent to the condition that the only solution to the linear matrix equations

$$C_{\mathcal{A}}\, \triangle V + C_{\mathcal{B}}\, \triangle V\, S + C_{\mathcal{B}} V\, \triangle S = 0, \quad W^{\mathsf{H}} \triangle V = 0, \tag{14}$$

is $(\triangle V, \triangle S) = (0, 0)$. Thus, to prove the statement of the theorem we need to show that (14) has a nonzero solution if and only if (10) has a nonzero solution in the homogeneous case $\triangle\mathbb{P} \equiv 0$.

Assume there exists $(\triangle X, \triangle S) \neq (0, 0)$ satisfying (10), i.e., $\mathbb{L}_P(\triangle X, \triangle S) = 0$ and $\mathbb{L}_V(\triangle X, \triangle S) = 0$. Define

$$\triangle V = \begin{bmatrix} \triangle X\, S^{\ell-1} + X\mathbb{D}S^{\ell-1}(\triangle S) \\ \vdots \\ \triangle X\, S^1 + X\mathbb{D}S^1(\triangle S) \\ \triangle X \end{bmatrix}.$$

Then, directly by definition, $\mathbb{L}_V(\triangle X, \triangle S) = 0$ implies $W^{\mathsf{H}} \triangle V = 0$. Moreover,

$$C_{\mathcal{A}} \triangle V + C_{\mathcal{B}} \triangle V\, S = \begin{bmatrix} \mathbb{P}(\triangle X, S) + \sum_{j=1}^{\ell-1} A_j X\, \mathbb{D}S^j(\triangle S) + A_\ell X\big(\mathbb{D}S^{\ell-1}(\triangle S)\big) S \\ X\big(\mathbb{D}S^{\ell-2}(\triangle S)\big) S - X\, \mathbb{D}S^{\ell-1}(\triangle S) \\ \vdots \\ XS - X\, \mathbb{D}S^1(\triangle S) \end{bmatrix}$$

By (13),

$$\big(\mathbb{D}S^{j-1}(\triangle S)\big) S - \mathbb{D}S^j(\triangle S) = -S^{j-1} \triangle S. \tag{15}$$

Together with $\mathbb{L}_P(\triangle X, \triangle S) = 0$, this shows

$$C_{\mathcal{A}} \triangle V + C_{\mathcal{B}} \triangle V\, S = \begin{bmatrix} -A_\ell X S^{\ell-1} \triangle S \\ -X S^{\ell-2} \triangle S \\ \vdots \\ -X \triangle S \end{bmatrix} = -C_{\mathcal{B}} V \triangle S,$$

and hence the constructed $(\triangle V, \triangle S)$ is a nontrivial solution of (14).

For the other direction, assume that there exists $(\triangle V, \triangle S) \neq (0, 0)$ satisfying (14). Partition $\triangle V = [\triangle X_{\ell-1}^{\mathsf{H}}, \ldots, \triangle X_0^{\mathsf{H}}]^{\mathsf{H}}$ with $\triangle X_j \in \mathbb{C}^{n\times k}$. Then the condition $C_{\mathcal{A}} \triangle V + C_{\mathcal{B}} \triangle V\, S + C_{\mathcal{B}} V \triangle S = 0$ implies

$$\sum_{j=1}^{\ell-1} A_j \triangle X_j + A_\ell \triangle X_{\ell-1}\, S + A_\ell X S^{\ell-1} \triangle S = 0, \tag{16}$$

$$-\triangle X_j + \triangle X_{j-1}\, S + X S^{j-1} \triangle S = 0, \qquad \text{for } j = 1, \ldots, \ell. \tag{17}$$

First, note that either $\triangle X_0 \neq 0$ or $\triangle S \neq 0$, since otherwise (17) implies $(\triangle V, \triangle S) = 0$. By induction, (17) combined with (15) yields

$$\triangle X_j = \triangle X_0\, S^j + X\, \mathbb{D}S^j(\triangle S). \tag{18}$$

Inserted into (16) and using (15) for $j = \ell$, this gives $\mathbb{L}_P(\triangle X_0, \triangle S) = 0$. Moreover, (18) immediately implies $\mathbb{L}_V(\triangle X_0, \triangle S) = 0$ from $W^{\mathsf{H}} \triangle V = 0$, which concludes the proof. □

3.2 First-order perturbation expansions

In the following, we use Theorem 7 to derive first-order perturbation expansions. The overall Fréchet derivative of the nonlinear equations (9) with respect to $(\hat{X}, \hat{S})$ evaluated at (X, S) is given by

$$\begin{array}{rcl} \mathbb{L} : \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k} & \to & \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k} \\ (\triangle X, \triangle S) & \mapsto & \big(\mathbb{L}_P(\triangle X, \triangle S), \mathbb{L}_V(\triangle X, \triangle S)\big), \end{array} \tag{19}$$

where $\mathbb{L}_P$ and $\mathbb{L}_V$ are defined as in (11)–(12). By Theorem 7, $\mathbb{L}$ is invertible for a simple invariant pair (X, S). By the implicit function theorem [26], there are uniquely determined analytic functions $f_X : U(0) \to \mathbb{C}^{n\times k}$ and $f_S : U(0) \to \mathbb{C}^{k\times k}$ such that

$$f_X(0) = X,\ f_S(0) = S, \quad f_X(\triangle P) = X + \triangle X,\ f_S(\triangle P) = S + \triangle S,$$

for all $\triangle P \in U(0)$ and some open neighborhood $U(0) \subset (\mathbb{C}^{n\times k})^{\ell+1}$ around zero. Moreover, the Fréchet derivatives of these functions satisfy

$$\big(\mathbb{D}f_X(\triangle P), \mathbb{D}f_S(\triangle P)\big) = -\mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big). \tag{20}$$

Defining

$$\|\triangle P\| := \big\|[E_0, E_1, \ldots, E_\ell]\big\|_F, \tag{21}$$

this shows that the perturbed polynomial $P+\triangle P$ has an invariant pair $(\hat{X}, \hat{S})$ close to (X, S), satisfying

$$(\hat{X}, \hat{S}) = (X, S) - \mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big) + \mathcal{O}(\|\triangle P\|^2), \tag{22}$$

where the addition of pairs is understood elementwise, under the assumption that the invariant pair (X, S) is simple. Note that the first-order correction term may contain components in the "direction" of (X, S). Since invariant pairs are only determined up to a basis transformation, there is a whole manifold $\mathcal{M}$ of invariant pairs generated by (X, S):

$$\mathcal{M} = \big\{(XT, T^{-1}ST) :\ T \in \mathbb{C}^{k\times k} \text{ invertible}\big\} \subset \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k}.$$

To assess the sensitivity of (X, S) under perturbations, it is sensible to neglect components of the error term $(\hat{X}, \hat{S}) - (X, S)$ that are contained in $\mathcal{M}$. In first-order, this can be achieved by considering the tangent space of $\mathcal{M}$ at (X, S),

$$T_{(X,S)}\mathcal{M} = \big\{(XM, SM - MS) :\ M \in \mathbb{C}^{k\times k}\big\}, \tag{23}$$

and projecting out components of $\mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big)$ contained in $T_{(X,S)}\mathcal{M}$. To summarize, we have the following result characterizing the first-order sensitivity of (X, S).

Theorem 8 *Let (X, S) be a simple invariant pair for a regular matrix polynomial P. For sufficiently small $\|\triangle P\|$ the perturbed polynomial $P + \triangle P$ has a simple invariant pair $(\tilde{X}, \tilde{S})$ satisfying*

$$(\tilde{X}, \tilde{S}) = (X, S) - (I - \mathsf{Proj}) \circ \mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big) + \mathcal{O}(\|\triangle P\|^2),$$

where Proj *is the orthogonal projector onto the tangent space $T_{(X,S)}\mathcal{M}$ defined in (23).*

Proof. By (22),

$$\begin{aligned}(\hat{X}, \hat{S}) - (X, S) &= -\mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big) + \mathcal{O}(\|\triangle P\|^2)\\ &= -\mathsf{Proj} \circ \mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big) - (I - \mathsf{Proj}) \circ \mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big) + \mathcal{O}(\|\triangle P\|^2).\end{aligned}$$

Setting $(\tilde{X}_0, \tilde{S}_0) := (\hat{X}, \hat{S}) - \mathsf{Proj} \circ \mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big)$ and defining $L_{P+\triangle P}$ similarly as L_P in (11), we obtain

$$\begin{aligned}(\mathbb{P} + \triangle\mathbb{P})(\tilde{X}_0, \tilde{S}_0) &= \underbrace{(\mathbb{P} + \triangle\mathbb{P})(\hat{X}, \hat{S})}_{=0} - \mathbb{L}_{P+\triangle P}\Big(\mathsf{Proj} \circ \mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big)\Big)\\ &= -\mathbb{L}_P\Big(\mathsf{Proj} \circ \mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big)\Big) + \mathcal{O}(\|\triangle P\|^2),\end{aligned}$$

Note that $\mathbb{P}$ is zero on $\mathcal{M}$ and hence its Jacobian $\mathbb{L}_P$ vanishes on $T_{(X,S)}\mathcal{M}$. In particular, $\mathbb{L}_P\big(\mathsf{Proj} \circ \mathbb{L}^{-1}\big(\triangle\mathbb{P}(X,S), 0\big)\big) = 0$, implying $(\mathbb{P} + \triangle\mathbb{P})(\tilde{X}_0, \tilde{S}_0) = \mathcal{O}(\|\triangle P\|^2)$. Therefore, for sufficiently small $\triangle P$ there exists an invariant pair $(\tilde{X}, \tilde{S})$ of $P + \triangle P$ such that $(\tilde{X}, \tilde{S}) - (\tilde{X}_0, \tilde{S}_0) = \mathcal{O}(\|\triangle P\|^2)$. Combined with the definition of $(\tilde{X}_0, \tilde{S}_0)$, this concludes the proof. □

We remark that Theorem 8 could be used to define a suitable condition number for an invariant pair (X, S) as the norm of $(I - \mathsf{Proj}) \circ \mathbb{L}^{-1}(\cdot, 0)$ induced by the norm (21).

3.3 The case $k = 1$

It is instructive to specialize the result of Theorem 8 to the case of eigenvectors, $k = 1$. In this case, $X \equiv x \in \mathbb{C}^n \setminus \{0\}$, $S \equiv \lambda \in \mathbb{C}$. Without loss of generality, we may assume $\|x\|_2 = 1$ and consider the fixed normalization vector $W = x$. Then the nonlinear matrix equations (9) amount to

$$(P + \triangle P)(\hat{\lambda}) \cdot \hat{x} = 0, \quad x^{\mathsf{H}}\hat{x} - 1 = 0.$$

The Fréchet derivative (19) with respect to $(\hat{x}, \hat{\lambda})$ evaluated at (x, λ) can be written in matrix form as

$$\mathbb{L} = \begin{bmatrix} P(\lambda) & P'(\lambda)x \\ x^{\mathsf{H}} & 0 \end{bmatrix}.$$

Theorem 7 states that $\mathbb{L}$ is invertible for a simple eigenvalue; which is in accordance with results from [2]. For $k = 1$, the tangent space $T_{(X,S)}\mathcal{M}$ defined in (23) and featuring prominently in Theorem 8 reduces to the one-dimensional linear space $\{(x\mu, 0) : \mu \in \mathbb{C}\}$ and hence the projector takes the form $\mathsf{Proj} = \left[\begin{smallmatrix} xx^{\mathsf{H}} & 0 \\ 0 & 0 \end{smallmatrix}\right]$. A straightforward calculation shows that $(I - \mathsf{Proj}) \circ \mathbb{L}^{-1}$ is given by

$$(I - \mathsf{Proj}) \circ \begin{bmatrix} X_\perp \big(Z_\perp^{\mathsf{H}} P(\lambda) X_\perp\big)^{-1} Z_\perp^{\mathsf{H}} & x \\ y^{\mathsf{H}}/(y^{\mathsf{H}}P'(\lambda)x) & 0 \end{bmatrix} = \begin{bmatrix} X_\perp \big(Z_\perp^{\mathsf{H}} P(\lambda) X_\perp\big)^{-1} Z_\perp^{\mathsf{H}} & 0 \\ y^{\mathsf{H}}/(y^{\mathsf{H}}P'(\lambda)x) & 0 \end{bmatrix}.$$

where the columns of $X_\perp, Z_\perp$ form orthonormal bases of $(\operatorname{span} x)^\perp$, $\operatorname{span}(P'(\lambda)x)^\perp$, respectively, and y denotes a normalized left eigenvector belonging to λ. Hence, Theorem 8 implies the perturbation expansions

$$\begin{aligned} \tilde{x} &= x - X_\perp \big(Z_\perp^{\mathsf{H}} P(\lambda) X_\perp\big)^{-1} Z_\perp^{\mathsf{H}} \triangle P(\lambda)x + \mathcal{O}(\|\triangle P\|^2), \\ \tilde{\lambda} &= \lambda - \frac{1}{y^{\mathsf{H}}P'(\lambda)x} y^{\mathsf{H}} \triangle P(\lambda)x + \mathcal{O}(\|\triangle P\|^2), \end{aligned}$$

which is again in accordance with results from [2, 5].

4 Computation via Linearization

In this section, we discuss the computation of invariant pairs of a matrix polynomial from invariant pairs of a corresponding linearization of the matrix polynomial. Solving polynomial eigenvalue problems by linearization is the most established method for computing eigenvalues and eigenvectors of polynomial eigenvalue problems of moderate size. Excellent introductions to numerical solvers for polynomial eigenvalue problems are given in the overview papers [39, 31].

In principle, it is possible to construct invariant pairs by combining several eigenvalue / eigenvector pairs. However, such a construction runs into conceptual and numerical difficulties as soon as some of the eigenvalues are (nearly) multiple. In contrast – as shown by the perturbation analysis in the previous section – invariant pairs remain well-posed objects in the presence of multiple eigenvalues as long as the algebraic eigenvalue multiplicities of the invariant pair match those of the matrix polynomial.

In Section 4.1 we discuss the relationship between invariant pairs of a matrix polynomial and the corresponding invariant pairs of its linearization in more detail. These results are put into practice in Section 4.2, where several strategies to extract an invariant pair of a matrix

polynomial from an invariant pair of its linearization are discussed. These are numerically tested in Section 4.3.

4.1 Linearization of Matrix Polynomials

The standard way to solve a polynomial eigenvalue problem (1) of degree $\ell \geq 2$ is to convert $P(\lambda)$ into a linear $\ell n \times \ell n$ pencil

$$L(\lambda) = \mathcal{A} + \lambda\mathcal{B}$$

having the same spectrum as $P(\lambda)$ and then solve this linear eigenvalue problem by a standard solver, e.g., the QZ algorithm [16, 25, 32]. A frequently used linearization is the companion form (5). This linearization has the property that

$$C_{\mathcal{B}} \boxplus\!\!\rightarrow C_{\mathcal{A}} := \begin{bmatrix} A_\ell & 0 & \cdots & 0 & 0 \\ 0 & I_n & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & I_n & 0 \end{bmatrix} + \begin{bmatrix} 0 & A_{\ell-1} & A_{\ell-2} & \cdots & A_0 \\ 0 & -I_n & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -I_n & 0 \end{bmatrix} = e_1 \otimes \begin{bmatrix} A_\ell & A_{\ell-1} & \ldots & A_0 \end{bmatrix}.$$

Here, following the notation introduced in [29], the so called *column shifted sum* $X \boxplus\!\!\rightarrow Y$ appends zero block columns to the right of the matrix X and to the left of the matrix Y and then adds up the enlarged matrices.

Using the column shifted sum it is possible to define a whole space of potential linearizations of P by

$$\mathcal{L}_1(P) = \left\{ \mathcal{A} + \lambda\mathcal{B} : \ \mathcal{B} \boxplus\!\!\rightarrow \mathcal{A} = v \otimes \begin{bmatrix} A_\ell & A_{\ell-1} & \ldots & A_0 \end{bmatrix}, \ v \in \mathbb{C}^\ell \right\}.$$

In [29] it was shown that almost all pencils in $\mathcal{L}_1(P)$ are linearizations of P. Furthermore, if $L(\lambda) = \mathcal{A} + \lambda\mathcal{B} \in \mathcal{L}_1(P)$ then

$$\mathcal{A} = \begin{bmatrix} W + \left(v \otimes \begin{bmatrix} A_{\ell-1} & \ldots & A_1 \end{bmatrix} \right), & v \otimes A_0 \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} v \otimes A_\ell, & -W \end{bmatrix},$$

where $W \in \mathbb{C}^{\ell n \times (\ell-1)n}$ is chosen arbitrarily [29, Theorem 3.5].

If (X, S) is an invariant pair for P then for any potential linearization $L(\lambda) = \mathcal{A} + \lambda\mathcal{B} \in \mathcal{L}_1(P)$ it holds that

$$\mathcal{A} \begin{bmatrix} XS^{\ell-1} \\ \vdots \\ X \end{bmatrix} + \mathcal{B} \begin{bmatrix} XS^{\ell-1} \\ \vdots \\ X \end{bmatrix} S = (\mathcal{B} \boxplus\!\!\rightarrow \mathcal{A}) \begin{bmatrix} XS^{\ell} \\ XS^{\ell-1} \\ \vdots \\ X \end{bmatrix} = v \otimes \begin{bmatrix} A_\ell & A_{\ell-1} & \ldots & A_0 \end{bmatrix} \begin{bmatrix} XS^{\ell} \\ XS^{\ell-1} \\ \vdots \\ X \end{bmatrix} = 0. \tag{24}$$

This generalizes (6) and shows that every invariant pair (X, S) of P can be used to construct an invariant pair of $L(\lambda) = \mathcal{A} + \lambda\mathcal{B}$. The converse question, whether an invariant pair (Y, S) of the linearization can be used to construct an invariant pair (X, S) of P, is answered in the following theorem. This is an extension of the eigenvector recovery property for $\mathcal{L}_1(P)$ shown in [29, Theorem 3.8].

Theorem 9 *Let $L(\lambda) = \mathcal{A} + \lambda\mathcal{B} \in \mathcal{L}_1(P)$ be a linearization of a regular matrix polynomial P. Then for every simple invariant pair $(Y, S) \in \mathbb{C}^{\ell n \times k} \times \mathbb{C}^{k \times k}$ of L there exists $X \in \mathbb{C}^{n \times k}$ such that $Y = V_\ell(X, S)$ and (X, S) is a simple invariant pair of P.*

Proof. An invariant pair (Y, S) of the matrix pencil $\mathcal{A} + \lambda\mathcal{B}$ satisfies

$$\mathcal{A}Y + \mathcal{B}YS = 0. \tag{25}$$

In the following, we consider S fixed and will show that the relation (25) implies $Y = V_\ell(X, S)$ for some $X \in \mathbb{C}^{n\times k}$. It then readily follows from (24) combined with $v \neq 0$ (otherwise, L would not be a linearization) and Lemma 6 that (X, S) is a simple invariant pair of P.

Let S have f mutually different eigenvalues $\lambda_1, \ldots, \lambda_f$ with algebraic multiplicities k_i partitioned into partial multiplicities $k_{i,1}, \ldots, k_{i,g_i}$, where g_i denotes the geometric multiplicity of λ_i. To classify all matrices Y satisfying (25) we first transform S to Jordan canonical form: $T^{-1}ST = J$ where T is invertible and $J = \operatorname{diag}\big(J_1, \ldots, J_f\big)$ with $J_i \in \mathbb{C}^{k_i\times k_i}$ containing the Jordan blocks for λ_i. Setting $\tilde{Y} = YT$, (25) becomes equivalent to

$$\mathcal{A}\tilde{Y} + \mathcal{B}\tilde{Y}J = 0. \tag{26}$$

Since (Y, S) is assumed to be simple, the partial eigenvalue multiplicities of $\mathcal{A}+\lambda\mathcal{B}$ match those of S and J. We can therefore choose $Y_i = \big[Y_{i,1}, \ldots, Y_{i,g_i}\big] \in \mathbb{C}^{\ell n\times k_i}$ such that $Y_{ij} \in \mathbb{C}^{\ell n\times k_{ij}}$ contains the jth Jordan chain of $\mathcal{A} + \lambda\mathcal{B}$ belonging to λ_i. A result by Košir [24, Theorem 4] implies that $\tilde{Y}$ satisfies (26) if and only if it takes the form

$$\tilde{Y} = \big[Y_1H_1,\ Y_2H_2,\ \ldots,\ Y_fH_f\big], \tag{27}$$

where $H_i \in \mathbb{C}^{k_i\times k_i}$ commutes with J_i (i.e., H_i is a block matrix partitioned conformally with J_i and each block is an upper triangular Toeplitz matrix [14, Pg. 221]). The discussion in Appendix A reveals that there is $X_i \in \mathbb{C}^{n\times k_i}$ such that $Y_i = V_\ell(X_i, J_i)$. Setting

$$\tilde{X} = \big[X_1,\ \ldots\ X_f\big], \quad H = \operatorname{diag}\big(H_1, \ldots, H_f\big)$$

the relation (27) can therefore be written as

$$\tilde{Y} = V_\ell(\tilde{X}, J)H = V_\ell\big(\tilde{X}H, J\big),$$

where we used the fact that J commutes with H. The proof is concluded by observing $Y = \tilde{Y}T^{-1} = V_\ell\big(\tilde{X}HT^{-1}, S\big)$ and setting $X = \tilde{X}HT^{-1}$. □

4.2 Extraction

In the following, we put the result of Theorem 9 into practice and discuss computational approaches to extracting an approximate invariant pair $(\tilde{X}, \tilde{S})$ for $P(\lambda)$ from a *computed* invariant pair $(\tilde{Y}, \tilde{S})$ of the linearization $L(\lambda)$.

Consider first the single vector case. Let $(\tilde{y}, \tilde{\lambda})$ be an approximate eigenpair for $L(\lambda) \in \mathcal{L}_1(P)$ and partition $\tilde{y} = \big[\tilde{y}_\ell^{\mathsf{H}} \quad \ldots \quad \tilde{y}_1^{\mathsf{H}}\big]^{\mathsf{H}}$ with $\tilde{y}_j \in \mathbb{C}^n$. In [20] it was shown for the companion linearization that a good choice for an approximate eigenvector of P is $\tilde{x} := \tilde{y}_\ell$ if $|\lambda| > 1$ and $\tilde{x} := \tilde{y}_1$ otherwise. The motivation behind this idea is that in exact arithmetic we have $y = \begin{bmatrix} \lambda^{\ell-1}x \\ \vdots \\ x \end{bmatrix}$ for an eigenvector x of P associated with λ. Hence, we can expect that – depending on the magnitude of λ – either the first or the last components of y will suffer least from cancellation in floating point arithmetic.

If (Y, S) is a simple invariant pair of $L(\lambda) \in \mathcal{L}_1(P)$ we can extend the ideas above and attempt to extract an invariant pair for $P(\lambda)$ from one of the block components $Y_j \in \mathbb{C}^{n\times k}$ of $Y = \begin{bmatrix} Y_\ell^{\mathsf{H}} & \dots & Y_1^{\mathsf{H}} \end{bmatrix}^{\mathsf{H}}$. In fact, for the block Y_1 of Y the feasibility of such an approach follows already from Theorem 9. For the other block components of Y the following lemma provides a necessary and sufficient condition.

Lemma 10 *Let (Y, S), $Y \in \mathbb{C}^{\ell n\times k}$, $S \in \mathbb{C}^{k\times k}$ be a simple invariant pair of $L(\lambda) \in \mathcal{L}_1(P)$ and let Y be partitioned as $Y = \begin{bmatrix} Y_\ell^{\mathsf{H}} & \dots & Y_1^{\mathsf{H}} \end{bmatrix}^{\mathsf{H}}$ with $Y_j \in \mathbb{C}^{n\times k}$, $j = 1, \dots, \ell$. Then, for any $j \in [2, \ell]$, (Y_j, S) is a simple invariant pair of $P(\lambda)$ if and only if S is nonsingular.*

Proof. Theorem 9 implies that (Y_1, S) is a simple invariant pair and $Y_j = Y_1 S^{j-1}$. We obtain

$$\mathbb{P}(Y_j, S) = A_\ell Y_j S^\ell + \cdots + A_1 Y_j S + A_0 Y_j = \mathbb{P}(Y_1, S) S^{j-1} = 0.$$

If S is nonsingular then this relation implies that (Y_j, S) is an invariant pair. Moreover,

$$\operatorname{rank}\left(V_\ell(Y_j, S)\right) = \operatorname{rank}\left(V_\ell(Y_1, S) S^{j-1}\right) \tag{28}$$

shows that (Y_j, S) is minimal and therefore a simple invariant pair. If S is singular then, by (28), (Y_j, S) is not minimal and is therefore not a simple invariant pair. ∎

Lemma 10 reveals that every block component of a *computed* simple and minimal invariant pair of $L(\lambda)$ is a candidate for approximating a simple invariant pair of $P(\lambda)$, provided that S is nonsingular. In the following we discuss four different strategies for extracting invariant pairs.

Extraction I (normwise) A heuristic choice for $\tilde{Y}_j$ is to choose the first block component of $\tilde{Y}$ if $\|S\| > 1$ and the last block component of $\tilde{Y}$ if $\|S\| < 1$. This is a direct generalization of the extraction strategy proposed in [20] for the single vector case.

Extraction II (polyeig) A more refined choice, inspired by the current extraction procedure in MATLAB's `polyeig`, is obtained by choosing j such that the residual

$$R(\tilde{Y}_j, \tilde{S}) := \frac{\left\|\mathbb{P}(\tilde{Y}_j, \tilde{S})\right\|_F}{\|\tilde{Y}_j\|_F} \tag{29}$$

is minimized.

Extraction III (GSVD) The above strategy can be further refined by minimizing among arbitrary $n \times k$ matrices. For a given $\tilde{S} \in \mathbb{C}^{k\times k}$ the optimal residual is obtained for $\tilde{X} \in \mathbb{C}^{n\times k}$ satisfying

$$R(\tilde{X}, \tilde{S}) = \min_{X \in \mathbb{C}^{n\times k}\setminus\{0\}} R(X, \tilde{S}),$$

with R defined as in (29). It follows that the vector $\operatorname{vec}(\tilde{X})$ is a right singular vector associated with the smallest singular value of the matrix $K := \sum_{j=0}^{\ell} \left(\tilde{S}^j\right)^{\mathrm{T}} \otimes A_j \in \mathbb{C}^{kn\times kn}$. However, the cost for solving this dense $kn \times kn$ SVD problem grows proportionally with k^3n^3 and is therefore not practicable for larger problems. To avoid this excessive computational cost, we therefore propose the following strategy based on the generalized singular value decomposition

(GSVD) [16]. An approximate minimizer of $R(\cdot, \tilde{S})$ can be obtained by restricting $\tilde{X}$ to be a linear combination of the block components of $\tilde{Y}$, that is

$$\tilde{X} = \gamma_1 \tilde{Y}_1 + \cdots + \gamma_\ell \tilde{Y}_\ell, \quad c = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_\ell \end{bmatrix} \in \mathbb{C}^\ell.$$

Since $\mathbb{P}(\tilde{X}, \tilde{S}) = \gamma_1 \mathbb{P}(\tilde{Y}_1, \tilde{S}) + \cdots + \gamma_\ell \mathbb{P}(\tilde{Y}_\ell, S)$ it follows that

$$\begin{aligned} R(\tilde{X}, \tilde{S}) &= \frac{\left\| \gamma_1 \mathbb{P}(\tilde{Y}_1, \tilde{S}) + \cdots + \gamma_\ell \mathbb{P}(\tilde{Y}_\ell, \tilde{S}) \right\|_F}{\left\| \gamma_1 \tilde{Y}_1 + \cdots + \gamma_\ell \tilde{Y}_\ell \right\|_F} \\ &= \frac{\left\| \left[\operatorname{vec}(\mathbb{P}(\tilde{Y}_1, \tilde{S})), \quad \ldots, \quad \operatorname{vec}(\mathbb{P}(\tilde{Y}_\ell, \tilde{S})) \right] c \right\|_2}{\left\| \left[\operatorname{vec}(\tilde{Y}_1), \quad \ldots \quad , \operatorname{vec}(\tilde{Y}_\ell) \right] c \right\|_2} =: \frac{\|Mc\|_2}{\|Nc\|_2}. \end{aligned}$$

Hence, the vector c that minimizes $R(\tilde{X}, \tilde{S})$ is the generalized singular vector associated with the smallest generalized singular value of the pair (M, N), where $M, N \in \mathbb{C}^{kn \times \ell}$ [16]. The cost of computing this vector is $O(kn\ell^2)$, which can be expected to remain small compared to the cost of computing the approximate invariant pair $(\tilde{Y}, \tilde{S})$ of $L(\lambda)$.

Extraction IV (structured) A rather different strategy to extract an approximate invariant pair $(\tilde{X}, \tilde{S})$ for P from $(\tilde{Y}, \tilde{S})$ is to consider structured projections of $\tilde{Y}$. In this approach, we choose $\tilde{X}$ as the solution to the minimization problem

$$\min_{\tilde{X} \in \mathbb{C}^{n \times k} \setminus \{0\}} \left\| \begin{bmatrix} \tilde{X} \tilde{S}^{\ell-1} \\ \vdots \\ \tilde{X} \end{bmatrix} - \begin{bmatrix} \tilde{Y}_\ell \\ \vdots \\ \tilde{Y}_1 \end{bmatrix} \right\|_F. \tag{30}$$

The following theorem provides an explicit solution to this problem.

Theorem 11 *The unique solution $X \in \mathbb{C}^{n \times k}$ that minimizes* (30) *is given by*

$$X = \left(\sum_{j=0}^{\ell-1} \tilde{Y}_{j+1} (\tilde{S}^j)^{\mathsf{H}} \right) \left(\sum_{j=0}^{\ell-1} \tilde{S}^j (\tilde{S}^j)^{\mathsf{H}} \right)^{-1}$$

Proof. Vectorizing (30) leads to the linear least-squares problem

$$\min_{x \in \mathbb{C}^{nk}} \left\| \begin{bmatrix} (\tilde{S}^{\ell-1})^{\mathsf{T}} \otimes I_n \\ \vdots \\ I_n \end{bmatrix} x - \begin{bmatrix} \operatorname{vec}(\tilde{Y}_\ell) \\ \vdots \\ \operatorname{vec}(\tilde{Y}_1) \end{bmatrix} \right\|_2,$$

The corresponding normal equations are given by

$$\sum_{j=0}^{\ell-1} \left((\tilde{S}^j)^{\mathsf{T}} \otimes I_n \right)^{\mathsf{H}} \left((\tilde{S}^{\ell-1})^{\mathsf{T}} \otimes I_n \right) x = \sum_{j=0}^{\ell-1} \left((\tilde{S}^j)^{\mathsf{T}} \otimes I_n \right)^{\mathsf{H}} \operatorname{vec}(\tilde{Y}_{j+1}),$$

leading to

$$\left[\left(\overline{\tilde{S}^{\ell-1}}(\tilde{S}^{\ell-1})^{\mathrm{T}} \otimes I_n\right) + \cdots + \left(\overline{\tilde{S}}\tilde{S}^{\mathrm{T}} \otimes I_n\right) + I_{kn}\right] x = \sum_{j=0}^{\ell-1} \left(\overline{\tilde{S}^{j}} \otimes I_n\right) \mathrm{vec}(\tilde{Y}_{j+1}).$$

Reformulation in terms of matrices gives

$$X\left[\tilde{S}^{\ell-1}(\tilde{S}^{\ell-1})^{\mathsf{H}} + \cdots + \tilde{S}\tilde{S}^{\mathsf{H}} + I_k\right] = \tilde{Y}_\ell(\tilde{S}^{\ell-1})^{\mathsf{H}} + \cdots + \tilde{Y}_2\tilde{S}^{\mathsf{H}} + \tilde{Y}_1.$$

Since the sum in the square brackets is positive definite and therefore nonsingular the result follows. □

4.3 Numerical comparison of the extraction strategies

It is immediately clear that the `polyeig` approach (**Extraction II**) is at least as good (in therms of the residual norm) as the normwise extraction (**Extraction I**) since it picks out the block of $\tilde{Y}$ that leads to the smallest residual. Also, we can expect that the GSVD approach (**Extraction III**) will perform at least as good or better than the `polyeig` approach since it tries to find a linear combination of all subblocks of $\tilde{Y}$ that minimizes the residual. Nevertheless, since the GSVD problem may be ill-conditioned it is not immediately clear in practice that this approach really leads to a numerically smaller residual. Also, it is unclear how the structured extraction performs in comparison as it does not aim to minimize the residual but rather projects the approximate invariant subspace $\tilde{Y}$ of the linearization onto a subspace of matrices with the right structural properties for an invariant subspace of a linearization.

To numerically compare the four different extraction strategies described above for a wide range of realistic problems we use the NLEVP collection of polynomial and nonlinear eigenvalue problems [7], from which we selected the 24 polynomial test problems with $n \leq 500$. For each test problem we extract invariant subspaces according to one of the following three criteria: (1) the 4 smallest eigenvalues in magnitude, (2) the 4 largest eigenvalues in magnitude, (3) the 2 smallest and 2 largest eigenvalues in magnitude.

To obtain these invariant subspace one could directly compute the corresponding eigenvalue/eigenvector pairs of the linearization and combine them to an invariant pair. However, as discussed earlier this may be numerically unstable. We rather want to avoid eigenvectors and work directly with invariant subspaces. To achieve this we first compute the Schur decomposition of the companion form (5) using MATLAB's `qz` function. This function returns unitary matrices Q and Z and upper triangular matrices $T_{\mathcal{A}}$ and $T_{\mathcal{B}}$, such that

$$QC_{\mathcal{A}}Z = T_{\mathcal{A}}, \quad QC_{\mathcal{B}}Z = T_{\mathcal{B}}.$$

Using the `ordqz` function in MATLAB this Schur decomposition is reordered such that the upper left 4×4 block of the pair $(T_{\mathcal{A}}, T_{\mathcal{B}})$ encodes the four smallest eigenvalues in magnitude for test case (1) and correspondingly the largest or largest/smallest eigenvalues for the other test cases. An invariant pair for the corresponding eigenvalues of the linearization is now given by $(Z_{(:,1:4)}, -T_{\mathcal{B}(1:4,1:4)}^{-1} T_{\mathcal{A}(1:4,1:4)})$ (MATLAB notation is used to denote submatrices). We then apply the different extraction strategies to obtain approximate invariant pairs $(\tilde{X}, -T_{\mathcal{B}(1:4,1:4)}^{-1} T_{\mathcal{A}(1:4,1:4)})$ of the original polynomial problem and measure the performance of the extraction strategies by comparing the residuals $R(\tilde{X}, -T_{\mathcal{B}(1:4,1:4)}^{-1} T_{\mathcal{A}(1:4,1:4)})$.

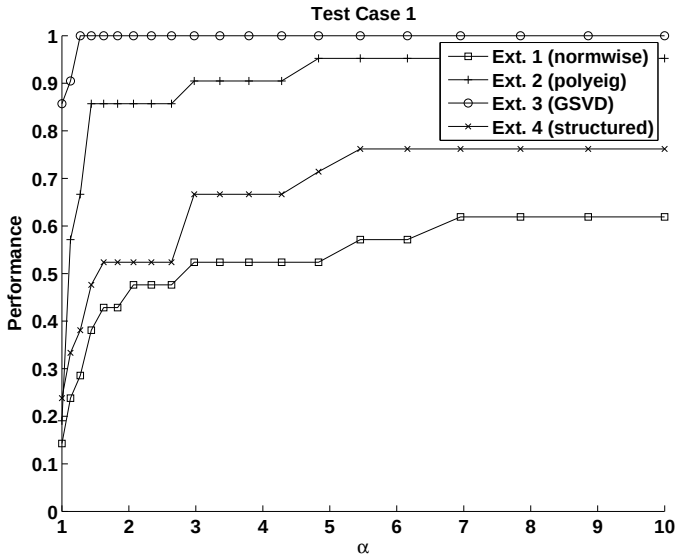

Figure 1: Performance diagram for the extraction of the smallest eigenvalues.

The results of the comparisons are presented in the form of performance diagrams in Figures 1,2 and 3. For a given factor α the performance is defined as the percentage of test cases for which the residual of the extracted invariant pair does not exceed α times the lowest residual achieved by any of the tested methods. In all three test cases the GSVD based extraction (**Extraction III**) turns out to be the method with the best performance. Since the additional cost of the GSVD computation is small compared to the solution of the overall polynomial eigenvalue problem, this strategy is therefore the one we recommend among the tested extraction methods. The normwise method (**Extraction I**) always performs worst and is therefore not recommended for the extraction of invariant subspaces. Remarkably, the structured approach (**Extraction IV**) performs reasonably well given that it does not attempt to achieve any residual minimization.

5 Refinement

In this section we discuss efficient iterative refinement strategies for approximate invariant pairs of a matrix polynomial P. Refinement is a crucial ingredient for the development of robust polynomial eigenvalue solvers that are based on linearizing the matrix polynomial P since these methods are not always backward stable [20]. Another interesting application arises in numerical continuation of eigenvalues for matrix polynomials as discussed by Beyn and Thümmler in [9].

5.1 Basic Algorithm

Given an approximation (X_0, S_0) to a simple invariant pair $(X, S) \in \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k}$ our aim is to compute a correction that brings (X_0, S_0) closer to (X, S). By Theorem 7, (X, S) is a regular value of the nonlinear matrix equations

$$\mathbb{P}(X, S) = 0, \quad \mathbb{V}(X, S) = 0, \tag{31}$$

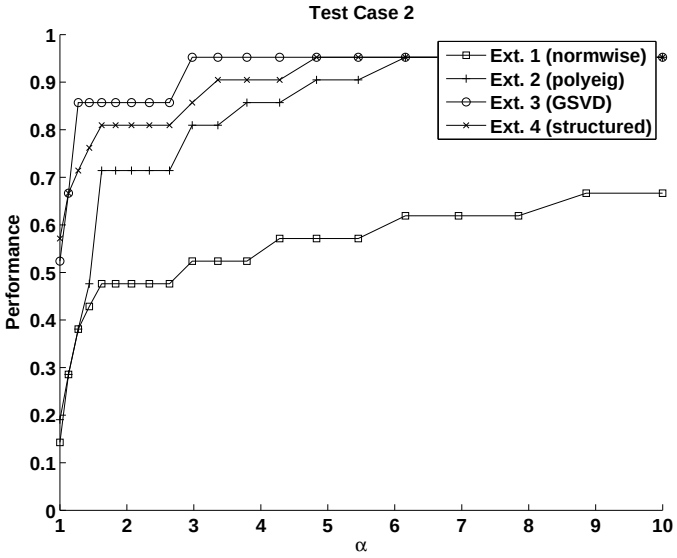

Figure 2: Performance diagram for the extraction of the largest eigenvalues.

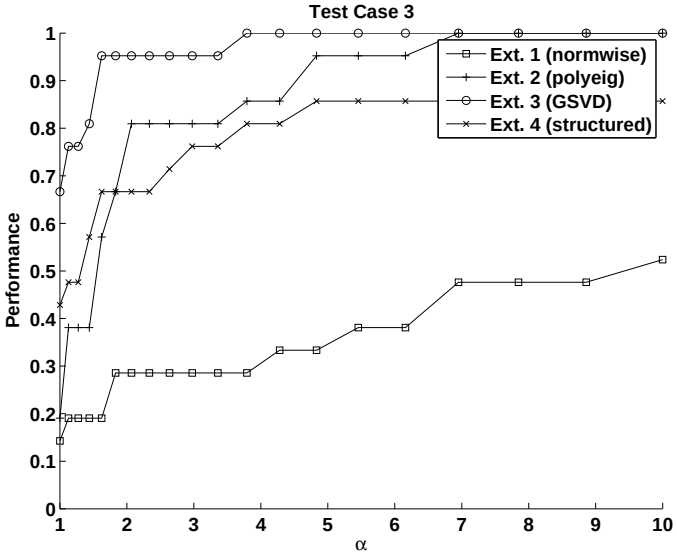

Figure 3: Performance diagram for the extraction of a block of small and large eigenvalues.

where $\mathbb{P}(X,S) = XA_0 + XA_1S + \cdots + XA_\ell S^\ell$ and $\mathbb{V}(X,S) = W^{\mathsf{H}}V_m(X,S) - I$ for some normalization matrix $W^{\mathsf{H}} = [W_{m-1}^{\mathsf{H}}, \ldots, W_0^{\mathsf{H}}] \in \mathbb{C}^{k\times mn}$. Newton's method applied to (31) with starting value (X_0, S_0) takes the form

$$(X_{p+1}, S_{p+1}) = (X_p, S_p) - \mathbb{L}_p^{-1}\big(\mathbb{P}(X_p, S_p), \mathbb{V}(X_p, S_p)\big), \tag{32}$$

where $\mathbb{L}_p$ is the Jacobian of $(\mathbb{P}, \mathbb{V})$ at the current iterate (X_p, S_p):

$$\mathbb{L}_p(\triangle X, \triangle S) = \left(\mathbb{P}(\triangle X, S_p) + \sum_{j=1}^{\ell} A_j X_p \,\mathbb{D}S_p^j(\triangle S), \sum_{j=0}^{m-1} W_j^{\mathsf{H}}\big(\triangle X\, S_p^j + X\,\mathbb{D}S_p^j(\triangle S)\big)\right),$$

see also (19). The invertibility of $\mathbb{L}_p$ and the local quadratic convergence of Newton's method is guaranteed by Theorem 7, provided of course that (X_0, S_0) is sufficiently close to (X, S).

In our implementation of (32) we keep the columns of $V_m(X_p, S_p)$ orthonormal and adapt W correspondingly in the course of the iteration. For this purpose, we compute a (compact) QR decomposition

$$V_m(X_p, S_p) = QR$$

with $Q \in \mathbb{C}^{mn\times k}$ such that $Q^{\mathsf{H}}Q = I$. It then follows directly that Q takes the form

$$Q = \begin{bmatrix} Q_0 R S_p^{m-1} R^{-1} \\ \vdots \\ Q_0 R S_p R^{-1} \\ Q_0 \end{bmatrix}.$$

for $Q_0 \in \mathbb{C}^{n\times k}$. Hence the replacement $(X_p, S_p) \leftarrow (Q_0, RS_pR^{-1})$ results in orthonormal $V_m(X_p, S_p)$. Moreover, by choosing $W = V_m(X_p, S_p)$ we have $\mathbb{V}(X_p, S_p) = 0$. Algorithm 1 summarizes the Newton method combined with this procedure.

Algorithm 1 Newton method for computing invariant pairs
Input: Initial pair $(X_0, S_0) \in \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k}$ such that $V_m(X_0, S_0)^{\mathsf{H}}V_m(X_0, S_0) = I_k$.
Output: Approximate solution (X_{p+1}, S_{p+1}) to (9).
1: $p \leftarrow 0$, $W \leftarrow V_m(X_0, S_0)$
2: **repeat**
3: Res $\leftarrow \mathbb{P}(X_p, S_p)$
4: Solve linear matrix equation $\mathbb{L}_p(\triangle X, \triangle S) = (\text{Res}, 0)$.
5: $\widetilde{X}_{p+1} \leftarrow X_p - \triangle X, \quad \widetilde{S}_{p+1} \leftarrow S_p - \triangle S$
6: Compute compact QR decomposition $V_m(X_{p+1}, S_{p+1}) = WR$.
7: $X_{p+1} \leftarrow \widetilde{X}_p R^{-1}, \quad S_{p+1} \leftarrow R\widetilde{S}_{p+1}R^{-1}$
8: **until** convergence

An extension of Algorithm 1 to nonlinear eigenvalue problems can be found in [28].

Remark 12 *The refinement procedure presented in Algorithm 1 is intended for initial pairs (X_0, S_0) that are already close to an invariant pair. This is, for example, the case for an inexact invariant pair obtained with any of the extraction procedures discussed in Section 4.2 in* finite-precision arithmetic*, provided that the invariant pair of interest is not too ill-conditioned.*

5.2 Solution of the correction equation

In the following, we discuss 3 approaches to solving the correction equation in Step 4 of Algorithm 1.

I. Kronecker products Vectorization and Kronecker products allow us to rewrite the linear matrix equation $\mathbb{L}_p(\triangle X, \triangle S) = (\mathrm{Res}, 0)$ as the $(nk + k^2) \times (nk + k^2)$ linear system

$$\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\triangle X) \\ \mathrm{vec}(\triangle S) \end{bmatrix} = \begin{bmatrix} \mathrm{vec}(\mathsf{Res}) \\ 0 \end{bmatrix}, \tag{33}$$

where

$$\begin{aligned} K_{11} &= \sum_{j=0}^{\ell} \big((S_p^j)^{\mathrm{T}} \otimes A_j\big), \quad & K_{12} &= \sum_{j=1}^{\ell} \big(I_k \otimes A_j X_p\big) K_{S_p^j}, \\ K_{21} &= \sum_{j=0}^{m-1} \big((S_p^j)^{\mathrm{T}} \otimes W_j^{\mathsf{H}}\big), \quad & K_{22} &= \sum_{j=1}^{m-1} \big(I_k \otimes W_j^{\mathsf{H}} X_p\big) K_{S_p^j}, \end{aligned}$$

with $K_{S_p^j}$ denoting the Kronecker product formulation of the Fréchet derivative $\mathbb{D}S_p^j$ (13):

$$K_{S_p^j} = \sum_{i=0}^{j-1} \big((S_p^{j-i-1})^{\mathrm{T}} \otimes S_p^i\big).$$

Solving (33) requires $\mathcal{O}((nk+k^2)^3)$ flops (floating point operations) and $\mathcal{O}((nk+k^2)^2)$ storage. This approach should therefore only be used for tiny values of k.

Remark 13 *For $k = 1$ and $m = 1$, the linear system (33) simplifies to*

$$\begin{bmatrix} P(\lambda) & P'(\lambda)x \\ W_0^{\mathsf{H}} & 0 \end{bmatrix} \begin{bmatrix} \triangle x \\ \triangle \lambda \end{bmatrix} = \begin{bmatrix} P(\lambda)x \\ 0 \end{bmatrix},$$

where we set $x \equiv X_p, \lambda \equiv S_p$.

II. Forward substitution By the Schur decomposition of S_p and an appropriate unitary transformation of (X_p, S_p), we may assume without loss of generality that S_p is in upper triangular form. The triangular structure of S_p allows to determine the columns of $\triangle X$ and $\triangle S$ successively in a forward substitution process. This was shown in [9] for quadratic eigenvalue problems and in [28] for nonlinear eigenvalue problems. We include the derivation of this forward substitution process for the sake of completeness, as it is needed in Approach III below.

In the following, we will drop the subscript p and simply write (X, S). The triangular structure of S implies that the equation $\mathbb{L}(\triangle X, \triangle S) = (\mathrm{Res}, 0)$ simplifies considerably for the first columns $\triangle x_1$ and $\triangle s_1$ of $\triangle X$ and $\triangle S$, respectively. In fact, it is not hard to see that

$$\begin{bmatrix} P(s_{11}) & \sum_{j=1}^{\ell} A_j X [\mathbb{D}S^j]_{11} \\ \sum_{j=0}^{m-1} s_{11}^j W_j^{\mathsf{H}} & \sum_{j=1}^{m-1} W_j^{\mathsf{H}} X [\mathbb{D}S^j]_{11} \end{bmatrix} \begin{bmatrix} \triangle x_1 \\ \triangle s_1 \end{bmatrix} = \begin{bmatrix} r_1 \\ 0 \end{bmatrix}, \tag{34}$$

where r_1 denotes the first column of Res and s_{11} is the first diagonal entry of S. The $k \times k$ matrix $[\mathbb{D}S^j]_{11}$ denotes the Fréchet derivative of the first column of S^j with respect to the first column of S. By (15), we have the recursion

$$\mathbb{D}S^1(\triangle S) = \triangle S, \quad \mathbb{D}S^j(\triangle S) = (\mathbb{D}S^{j-1}(\triangle S))S + S^{j-1}\triangle S, \quad j \geq 2,$$

implying

$$[\mathbb{D}S^1]_{11} = I_k, \quad [\mathbb{D}S^j]_{11} = s_{11}[\mathbb{D}S^{j-1}]_{11} + S^{j-1}, \quad j \geq 2. \tag{35}$$

Besides providing an efficient means for computing $[\mathbb{D}S^j]_{11}$, this also shows that $[\mathbb{D}S^j]_{11}$ is upper triangular.

Similar to the forward substitution process for solving lower triangular systems, we can derive an equation of the form (34) also for the second columns of $\triangle X$ and $\triangle S$, provided that the right hand side is updated accordingly. To describe this update, partition

$$\triangle X = [\triangle x_1, \triangle X_2], \quad \triangle S = [\triangle s_1, \triangle S_2], \quad \mathsf{Res} = [r_1, \mathsf{Res}_2],$$

and

$$S = \begin{bmatrix} s_{11} & s_{12} \\ 0 & S_{22} \end{bmatrix}, \quad S^j = \begin{bmatrix} s_{11}^j & [S^j]_{12} \\ 0 & S_{22}^j \end{bmatrix}.$$

Inserted into $\mathbb{L}(\triangle X, \triangle S) = (\mathrm{Res}, 0)$, we obtain the following linear matrix equation for the pair $(\triangle X_2, \triangle S_2) \in \mathbb{C}^{n\times(k-1)} \times \mathbb{C}^{k\times(k-1)}$:

$$\mathbb{P}(\triangle X_2, S_{22}) + \sum_{j=0}^{\ell} A_j X\, \mathbb{D}S^j\big([0, \triangle S_2]\big) \begin{bmatrix} 0 \\ I_{k-1} \end{bmatrix} = \widetilde{\mathsf{Res}}_2, \tag{36}$$

$$\sum_{j=0}^{m-1} W_j^{\mathsf{H}} \left(\triangle X_2 S_{22}^j + X\, \mathbb{D}S^j\big([0, \triangle S_2]\big) \begin{bmatrix} 0 \\ I_{k-1} \end{bmatrix} \right) = \widetilde{\mathsf{Ort}}_2. \tag{37}$$

with updated right-hand sides

$$\begin{aligned} \widetilde{\mathsf{Res}}_2 &:= \mathsf{Res}_2 - \sum_{j=0}^{\ell} A_j \left(\triangle x_1 [S^j]_{12} + X \mathbb{D}S^j([\triangle s_1, 0]) \begin{bmatrix} 0 \\ I_{k-1} \end{bmatrix} \right), \\ \widetilde{\mathsf{Ort}}_2 &:= -\sum_{j=1}^{m-1} W_j^{\mathsf{H}} \left(\triangle x_1 [S^j]_{12} + X\, \mathbb{D}S^j\big([\triangle s_1, 0]\big) \begin{bmatrix} 0 \\ I_{k-1} \end{bmatrix} \right). \end{aligned}$$

Letting r_2 and q_2 denote the first columns of $\widetilde{\mathsf{Res}}_2$ and $\widetilde{\mathsf{Ort}}_2$, respectively, this shows that the second columns $\triangle x_2, \triangle s_2$ of $\triangle X, \triangle S$ satisfy the linear system

$$\begin{bmatrix} P(s_{22}) & \sum_{j=1}^{\ell} A_j X [\mathbb{D}S^j]_{22} \\ \sum_{j=0}^{m-1} s_{22}^j W_j^{\mathsf{H}} & \sum_{j=1}^{m-1} W_j^{\mathsf{H}} X [\mathbb{D}S^j]_{22} \end{bmatrix} \begin{bmatrix} \triangle x_2 \\ \triangle s_2 \end{bmatrix} = \begin{bmatrix} r_2 \\ q_2 \end{bmatrix}, \tag{38}$$

where s_{22} denotes the first diagonal element of S_{22} and $[\mathbb{D}S^j]_{22}$ satisfies the recursion (35) with s_{11} replaced by s_{22}.

The described process can be continued in an analogous manner to compute all columns of $\triangle X$ and $\triangle S$. The cost of the overall algorithm is dominated by the solution of k linear systems of the form (34) and (38). Since each of these systems has order $n+k$, the overall cost is $\mathcal{O}(k(n+k)^3)$ flops, which compares favorably with the $\mathcal{O}((nk+k^2)^3)$ flops needed by the Kronecker product formulation. If the coefficients A_j of the matrix polynomial are sparse then (34) is a bordered sparse system and a sparse direct solver for bordered matrices [4] could be used. Moreover, it might be possible to extend ideas on Krylov subspace methods for parametrized systems [33] to design a Krylov subspace method that handles the k systems of the form (34), (38) for $s_{11}, \ldots, s_{kk}$ simultaneously.

III. Linearization Given a matrix polynomial P, the efficient solution of linear systems of the form $P(s)x = b$ for many different parameters $s \in \mathbb{C}$ and right-hand sides b by means of linearizing P has been discussed in [17, 33]. In the following, we extend these ideas to solve bordered systems of the form

$$\begin{bmatrix} P(s) & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{39}$$

for many different values of $s \in C$. The border matrices $A_{12} \in \mathbb{C}^{n\times k}$, $A_{21} \in \mathbb{C}^{k\times n}$, $A_{22} \in \mathbb{C}^{k\times k}$, and the right-hand side are different for each s, in some non-specified fashion.

Given a linearization $\mathcal{A} + \lambda\mathcal{B} \in \mathbb{L}_1(P)$, we have

$$(\mathcal{A} + s\mathcal{B})V_\ell(x_1, s) = v \otimes P(s)x_1$$

for arbitrary $s \in \mathbb{C}$, $x_1 \in \mathbb{C}^n$, and some fixed nonzero vector $v \in \mathbb{C}^\ell$ describing the linerization [29]. Note that $v \otimes P(s)x = v \otimes b$ if and only if $P(s)x = b$. This allows us to rewrite (39) as

$$\begin{bmatrix} \mathcal{A} + s\mathcal{B} & v \otimes A_{12} \\ w^{\mathsf{H}} \otimes A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} V_\ell(x_1, s) \\ x_2 \end{bmatrix} = \begin{bmatrix} v \otimes b_1 \\ b_2 \end{bmatrix} \tag{40}$$

where $w \in \mathbb{C}^\ell$ is any vector satisfying $[s^{\ell-1}, \ldots, s, 1]^{\mathsf{H}} w = 1$. Once the solution $\tilde{y} \in \mathbb{C}^{\ell n+k}$ to (40) is computed, we can extract x_2 from its trailing k entries and x_1 from its leading ℓn entries using any of the extraction strategies discussed in Section 4.2. Note that the conditioning for (40) might be significantly worse than for (39), but a full discussion of this effect is behind the scope of this paper. Instead, we refer to [17] for a related discussion and remark that there is no need to solve (40) very precisely thanks to the forgivingness of the outer Newton iteration [38].

To solve (40) efficiently for many different s we first compute a generalized Schur decomposition

$$Q^{\mathsf{H}}(\mathcal{A} + \lambda\mathcal{B})Z = T_{\mathcal{A}} + \lambda T_{\mathcal{B}} \tag{41}$$

with unitary matrices $Q, Z \in \mathbb{C}^{\ell n\times \ell n}$ and upper triangular matrices $T_{\mathcal{A}}, T_{\mathcal{B}} \in \mathbb{C}^{\ell n\times \ell n}$. Note that if the initial approximation (X_0, S_0) to the invariant pair was computed by solving the linearized eigenvalue problem combined with one of the extraction methods described in Section 4 then this decomposition is usually readily available. Setting

$$\tilde{A}_{12} = Q^{\mathsf{H}}\big(v \otimes A_{12}\big),\ \tilde{A}_{21} = \big(w^{\mathsf{H}} \otimes A_{21}\big)Z,\ \tilde{x}_1 = Z^{\mathsf{H}}V_\ell(x_1, s),\ \tilde{b}_1 = Q^{\mathsf{H}}(v \otimes b_1), \tag{42}$$

the linear system (40) becomes equivalent to

$$\begin{bmatrix} T_{\mathcal{A}} + sT_{\mathcal{B}} & \tilde{A}_{12} \\ \tilde{A}_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \tilde{b}_1 \\ b_2 \end{bmatrix}, \tag{43}$$

which is a bordered triangular system and can be solved, e.g., via a slightly modified QR or LU decomposition [10] that takes the structure into account. This requires $\mathcal{O}(k(\ell n + k)^2)$ flops for computing the decomposition and $\mathcal{O}((\ell n + k)^2)$ flops for solving the resulting upper triangular system. Setting up the transformed system (43) requires another $\mathcal{O}(k(\ell n)^2)$ flops.

In total, the overall cost of this approach for refining an invariant pair is $\mathcal{O}(\ell^3 n^3)$ flops for computing the generalized Schur decomposition (which needs to be performed only once throughout the entire Newton iteration or might already be available) plus $\mathcal{O}(k^2(\ell n + k)^2)$ flops for solving the k linear systems. This compares well with the $\mathcal{O}(k(n+k)^3)$ flops needed by the second approach, provided that ℓ stays small and n is sufficiently large.

Performance comparison To gain insight into the actual performance of the three approaches we measured the execution times needed for

- solving the linear system (33) in Approach I;
- solving k linear systems of the form (38) in Approach II;
- setting up (42) once and solving k linear systems of the form (43) in Approach III.

We run experiments in MATLAB 7.5 on a 2.20 GHz Intel Core2 Duo CPU with 2 GiB RAM. All approaches have been implemented in MATLAB in a rather straightforward fashion, with the exception that we have used a MEX interface to a slightly modified variant of the LAPACK routine ZGETRF for computing the LU decomposition of (43) within $\mathcal{O}(k(\ell n+k)^2)$ flops. The following two tables contain the obtained execution times in seconds for $n = 500, 1000, 2000$:

$n = 500$

k	I	II	III
2	0.57	0.17	0.14
4	3.9	0.33	0.28
32	∞	3.2	2.7
128	∞	20	23

$n = 1000$

k	I	II	III
2	3.8	1.11	0.56
4	28	2.24	1.16
32	∞	19	11
128	∞	97	79

$n = 2000$

k	I	II	III
2	28	7.7	2.4
4	∞	15	5.1
32	∞	126	44
128	∞	582	303

An entry ∞ indicates an out of memory error. As expected, Approach I is rather expensive and should only be used for tiny k and n. With the exception of $n = 500, k = 128$, Approach III is always faster than Approach II. However, it is important to note that these figures assume the availability of a Schur decomposition for the linearization. If this decomposition is not available (because, for example, the initial approximation to the invariant pair has been obtained by some other means), Approach III becomes much less attractive. The current implementation [32] of the QZ algorithm requires about 160 seconds for $n = 500$ and about 1450 seconds for $n = 1000$. Even taking into account that the new implementation of the QZ algorithm described in [25] (which is not yet included in MATLAB) may reduce these numbers by a factor $4 - 8$ it would require an excessive number of iterations to make Approach III competitive.

6 Numerical Examples

In this section, we illustrate the use of the presented concepts for two examples from [39, 7].

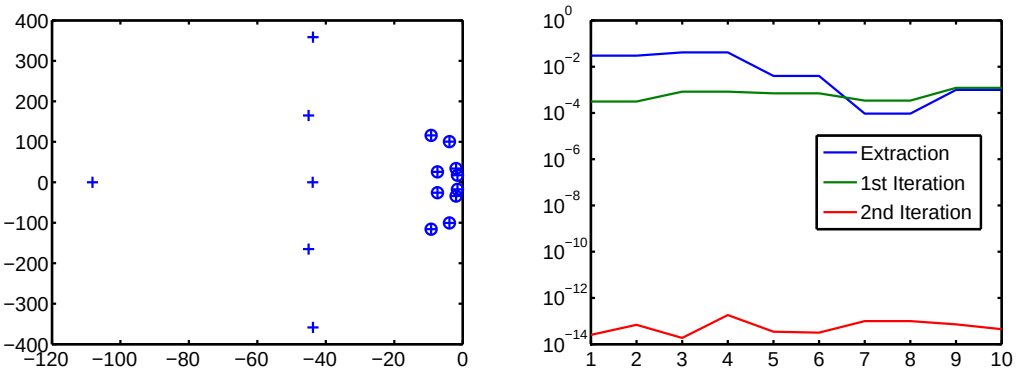

Figure 4: Power plant example from [7]. *Left plot:* Location of eigenvalues (crosses) and selected eigenvalues (circles). *Right plot:* Absolute error of the 10 selected eigenvalues after (i) extraction from the linearization, (ii) 1 Newton iteration, (iii) 2 Newton iterations.

Example 14 *The simplified dynamical model of a nuclear power plant from [7] leads to an* 8×8 *quadratic matrix polynomial that has been noted [39] to have rather ill-conditioned eigenvalues, mainly due to the bad scaling of the coefficient matrices. Using Extraction III based on the GSVD, we compute the invariant pair for the* 10 *rightmost eigenvalues from the linearization. "Exact eigenvalues" are obtained from a high precision arithmetic computation. As shown in Figure 4, the computed eigenvalues are rather inaccurate, with absolute errors of order* 10^{-2} *to* 10^{-4}*. Two Newton iterations applied to the extracted invariant pair reduce these errors down to almost machine precision. This indicates that iterative refinement for invariant pairs cures the effects of bad scaling, similarly as for linear systems [18].*

Example 15 *In [39] the following matrix polynomial was discussed:*

$$P(\lambda) = \lambda^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \lambda \begin{bmatrix} -2 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It has an eigenvalue $\lambda = 1$ *with algebraic multiplicity* 3*. A corresponding invariant pair is given by*

$$X = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

We perturb X *and* S *by setting* $X(3,3) = 1$ *and* $S(3,2) = 10^{-8}$*. The perturbed matrix* $\tilde{S}$ *has eigenvalues* $\lambda_1 = 1$*,* $\lambda_2 \approx 1.0001$*,* $\lambda_3 \approx 0.9999$*. The initial residual of* $(\tilde{X}, \tilde{S})$ *is* $R(\tilde{X}, \tilde{S}) \approx 0.73$*. After three refinement steps using approach II (Forward Substitution) we have* $R(\tilde{X}, \tilde{S}) \approx 3.89 \cdot 10^{-16}$*. The refined eigenvalues of* $\tilde{S}$ *are*

$$\begin{aligned} \tilde{\lambda}_1 &= 1, \\ \tilde{\lambda}_2 &= 0.999999999945053 + 7.654628153552778 \times 10^{-9}\mathrm{i}, \\ \tilde{\lambda}_3 &= 1.000000000005495 - 7.654628059339143 \times 10^{-9}\mathrm{i}. \end{aligned}$$

The refined matrix $\tilde{X}$ is given by (displayed to three decimal digits accuracy)

$$\tilde{X} = \begin{bmatrix} 0 & 7.07 \cdot 10^{-1} - 5.79 \cdot 10^{-4} i & 7.07 \cdot 10^{-1} + 1.10 \cdot 10^{-5} i \\ 7.07 \cdot 10^{-1} & 1.27 \cdot 10^{-17} - 1.04 \cdot 10^{-20} i & -2.98 \cdot 10^{-17} - 4.65 \cdot 10^{-22} i \\ 0 & 5.91 \cdot 10^{-36} + 2.25 \cdot 10^{-35} i & 6.75 \cdot 10^{-32} + 8.14 \cdot 10^{-32} i \end{bmatrix}.$$

The third row is close to zero and hence the span of this matrix is almost identical to the span of the original matrix X demonstrating that the invariant pair (X, S) was very well recovered even though S contains a Jordan block.

7 Conclusions

One aim of this paper is to promote the concept of invariant pairs for polynomial eigenvalue problems as a suitable way of handling several eigenvalues simultaneously. Several theoretical results, algorithms, and numerical experiments have been presented to support this concept. The benefits of using invariant pairs in applications are not fully explored yet. The experiments in Section 6 suggest that extracting and refining invariant pairs might have a positive impact on the accuracy in any polynomial eigenvalue computation. Also, we believe that invariant pairs can be a useful framework in the design and analysis of Krylov subspace and Jacobi-Davidson methods for solving large-scale polynomial eigenvalue problems [3, 23, 30]. Finally, we remark that some of the results presented in this paper can be extended to genuinely nonlinear eigenvalue problems [28].

References

[1] B. Adhikari, R. Alam, and D. Kressner. Structured eigenvalue condition numbers and linearizations for matrix polynomials. Technical report 2009-01, Seminar for applied mathematics, ETH Zurich, January 2009.

[2] A. L. Andrew, K.-W. E. Chu, and P. Lancaster. Derivatives of eigenvalues and eigenvectors of matrix functions. *SIAM J. Matrix Anal. Appl.*, 14(4):903–926, 1993.

[3] Z. Bai and Y. Su. SOAR: a second-order Arnoldi method for the solution of the quadratic eigenvalue problem. *SIAM J. Matrix Anal. Appl.*, 26(3):640–659 (electronic), 2005.

[4] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numer.*, 14:1–137, 2005.

[5] M. Berhanu. *The Polynomial Eigenvalue Problem.* PhD thesis, School of Mathematics, The University of Manchester, 2005.

[6] T. Betcke. Optimal scaling of generalized and polynomial eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 30(4):1320–1338, 2008.

[7] T. Betcke, N. Higham, V. Mehrmann, C. Schröder, and F. Tisseur. NLEVP: A collection of nonlinear eigenvalue problems. Technical Report 2008.40, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2008.

[8] 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.

[9] W.-J. Beyn and V. Thümmler. Continuation of invariant subspaces for parameterized quadratic eigenvalue problems. Technical report, University of Bielefeld, Department of Mathematics, 2008.

[10] Å. Björck. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, PA, 1996.

[11] J.-P. Dedieu and F. Tisseur. Perturbation theory for homogeneous polynomial eigenvalue problems. *Linear Algebra Appl.*, 358:71–94, 2003.

[12] J. E. Dennis, Jr., J. F. Traub, and R. P. Weber. The algebraic theory of matrix polynomials. *SIAM J. Numer. Anal.*, 13(6):831–845, 1976.

[13] H.-Y. Fan, W.-W. Lin, and P. Van Dooren. Normwise scaling of second order polynomial matrices. *SIAM J. Matrix Anal. Appl.*, 26(1):252–256, 2004.

[14] F.R. Gantmacher. *The Theory of Matrices*. Chelsea, New York, 1960.

[15] I. Gohberg, P. Lancaster, and L. Rodman. *Matrix polynomials*. Academic Press Inc., New York, 1982. Computer Science and Applied Mathematics.

[16] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[17] L. Grammont, N. J. Higham, and F. Tisseur. A framework for analyzing nonlinear eigenproblems and parametrized linear systems, 2009. Technical Report.

[18] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, PA, second edition, 2002.

[19] N. J. Higham and H.-M. Kim. Numerical analysis of a quadratic matrix equation. *IMA J. Numer. Anal.*, 20(4):499–519, 2000.

[20] N. J. Higham, R. C. Li, and F. Tisseur. Backward error of polynomial eigenproblems solved by linearization. *SIAM J. Matrix Anal. Appl.*, 29:1218–1241, 2007.

[21] N. J. Higham, D. S. Mackey, and F. Tisseur. The conditioning of linearizations of matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 28(4):1005–1028, 2006.

[22] N. J. Higham, D. S. Mackey, F. Tisseur, and S. D. Garvey. Scaling, sensitivity and stability in the numerical solution of quadratic eigenvalue problems. *Internat. J. Numer. Methods Engrg.*, 73(3):344–360, 2008.

[23] L. Hoffnung, R.-C. Li, and Q. Ye. Krylov type subspace methods for matrix polynomials. *Linear Algebra Appl.*, 415(1):52–81, 2006.

[24] T. Košir. Kronecker bases for linear matrix equations, with application to two-parameter eigenvalue problems. *Linear Algebra Appl.*, 249:259–288, 1996.

[25] B. Kågström and D. Kressner. Multishift variants of the QZ algorithm with aggressive early deflation. *SIAM J. Matrix Anal. Appl.*, 29(1):199–227, 2006.

[26] S. G. Krantz. *Function Theory of Several Complex Variables*. John Wiley & Sons Inc., New York, 1982.

[27] D. Kressner. *Numerical Methods for General and Structured Eigenvalue Problems*, volume 46 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2005.

[28] D. Kressner. A block Newton method for nonlinear eigenvalue problems. *Numer. Math.*, 114(2):355–372, 2009.

[29] 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.

[30] K. Meerbergen. Locking and restarting quadratic eigenvalue solvers. *SIAM J. Sci. Comput.*, 22(5):1814–1839 (electronic), 2000.

[31] V. Mehrmann and H. Voss. Nonlinear eigenvalue problems: A challenge for modern eigenvalue methods. *Mitt. der Ges. f. An. Mathematik und Mechanik*, 27:121–151, 2005.

[32] C. B. Moler and G. W. Stewart. An algorithm for generalized matrix eigenvalue problems. *SIAM J. Numer. Anal.*, 10:241–256, 1973.

[33] V. Simoncini and F. Perotti. On the numerical solution of $(\lambda^2 A + \lambda B + C)x = b$ and application to structural dynamics. *SIAM J. Sci. Comput.*, 23(6):1875–1897 (electronic), 2002.

[34] G. W. Stewart. Error bounds for approximate invariant subspaces of closed linear operators. *SIAM J. Numer. Anal.*, 8:796–808, 1971.

[35] G. W. Stewart. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Rev.*, 15:727–764, 1973.

[36] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, New York, 1990.

[37] J.-G. Sun. Perturbation expansions for invariant subspaces. *Linear Algebra Appl.*, 153:85–97, 1991.

[38] F. Tisseur. Newton's method in floating point arithmetic and iterative refinement of generalized eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 22(4):1038–1057, 2001.

[39] F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Rev.*, 43(2):235–286, 2001.

A Construction of Jordan chains

In the following we demonstrate how Jordan chains of a regular matrix polynomial P can be turned into Jordan chains of a linearization $\mathcal{A} + \lambda\mathcal{B} \in \mathcal{L}_1(P)$. This result is needed in the proof of Theorem 9 and the construction is rather similar to the ones given in [15, 39].

Let λ be a finite eigenvalue of P and consider an arbitrary vector $x \in \mathbb{C}^n$. Then

$$(\mathcal{A} + \lambda\mathcal{B})V_\ell(x, \lambda) = v \otimes P(\lambda)x \tag{44}$$

see [29] or the relation (24) for $k = 1$. Differentiating (44) with respect to λ yields

$$(\mathcal{A} + \lambda\mathcal{B})V_\ell'(x, \lambda) = v \otimes P'(\lambda)x - BV_\ell(x, \lambda), \tag{45}$$

where

$$V_\ell'(x, \lambda) = \begin{bmatrix} (l-1)\lambda^{l-2}x \\ \vdots \\ 2\lambda x \\ x \\ 0 \end{bmatrix}.$$

Chains of length 2***:*** Let us first consider a Jordan chain $x_1, x_2 \in \mathbb{C}^n$ of length 2 for P:

$$P(\lambda)x_1 = 0, \quad P(\lambda)x_2 + P'(\lambda)x_1 = 0.$$

Set

$$y_1 := V_\ell(x_1, \lambda), \quad y_2 := V_\ell(x_2, \lambda) + V_\ell'(x_1, \lambda).$$

Then (44) yields $(\mathcal{A} + \lambda\mathcal{B})y_1 = 0$ and (45) yields

$$(\mathcal{A} + \lambda\mathcal{B})y_2 = v \otimes P(\lambda)x_2 + v \otimes P'(\lambda)x_1 - By_1 = -By_1.$$

This shows that y_1, y_2 is a Jordan chain for $(\mathcal{A} + \lambda\mathcal{B})$. Note that we can write

$$[y_1, y_2] = V_\ell\left([x_1, x_2], \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}\right).$$

Chains of arbitrary length: Let us now consider a Jordan chain $x_1, \ldots, x_k \in \mathbb{C}^n$ of length k for P:

$$\sum_{i=1}^{j} \frac{1}{(i-1)!} P^{(i-1)}(\lambda)x_{j-i+1} = 0, \quad \text{for } j = 1, \ldots, k.$$

Set

$$y_j := \sum_{i=1}^{j} \frac{1}{(i-1)!} V_\ell^{(i-1)}(x_{j-i+1}, \lambda).$$

Repeated differentiation of (45) gives

$$(\mathcal{A} + \lambda\mathcal{B})V_\ell^{(i-1)}(x, \lambda) = v \otimes P^{(i-1)}(\lambda)x - (i-1)BV_\ell^{(i-2)}(x, \lambda)$$

and hence

$$\begin{aligned}(\mathcal{A}+\lambda\mathcal{B})y_j &= \sum_{i=1}^{j}\frac{1}{(i-1)!}(\mathcal{A}+\lambda\mathcal{B})V_\ell^{(i-1)}(x_{j-i+1},\lambda)\\ &= \sum_{i=1}^{j}\Big(v\otimes\frac{1}{(i-1)!}P^{(i-1)}(\lambda)x-\frac{(i-1)}{(i-1)!}BV_\ell^{(i-2)}(x,\lambda)\Big)\\ &= -\sum_{i=1}^{j}\frac{1}{(i-2)!}BV_\ell^{(i-2)}(x,\lambda)=-By_{j-1}.\end{aligned}$$

This shows that $y_1,\ldots,y_k$ is a Jordan chain for $(\mathcal{A}+\lambda\mathcal{B})$. Moreover, we can write

$$[y_1,\ldots,y_k]=V_\ell\big([x_1,\ldots,x_k],J_k(\lambda)\big),$$

where $J_k(\lambda)$ is a $k\times k$ Jordan block belonging to λ.