School of Mathematics, Meteorology
and Physics

Department of Mathematics

Preprint MPS_2009-05

10 July 2009

For linear eigenvalue problems, it is well known that working with invariant subspaces instead of eigenvectors o ers 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 all copies of the eigenvalue are included in the subspace. 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. As the focus of this paper is on numerical aspects, we shall not discuss this connection in more detail.

For $k = n$, any matrix S satisfying (2) is called a solvent. 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 treat the matrices X and S not as independent 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

$\alpha\beta^{\ell-1}A_1 + \alpha^2\beta^{\ell-2}A_2 + \cdots + \alpha^{\ell}A_{\ell}$ *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 [34, 35]. In this work, we will refrain from using such a homogenous formulation as it would significantly increase the level of technicality. Moreover, one of the advantages of deflating subspaces, their direct connection to the factors of the generalized Schur form, is lost when going to* $\ell > 1$*. Infinite eigenvalues contained in an invariant subspace can still be*

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

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

has full column rank. The smallest such $\mathbf{m}$ *is called minimality index of* $(\mathbf{X},\mathbf{S})$.

By the Cayley-Hamilton theorem, the minimality index of a minimal pair cannot exceed $\mathbf{k}$, 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 most of the discussion in this paper to minimal invariant pairs.

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

$$\operatorname{span} \mathbf{V}_\ell(\tilde{\mathbf{X}},\tilde{\mathbf{S}}) = \operatorname{span} \mathbf{V}_\ell(\mathbf{X},\mathbf{S}),$$

with $\mathbf{V}_\ell(\mathbf{X},\mathbf{S})$ *and* $\mathbf{V}_\ell(\tilde{\mathbf{X}},\tilde{\mathbf{S}})$ *defined as in (4).*

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

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

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

$$\mathbf{X}\mathbf{S} = \big[\tilde{\mathbf{X}}\tilde{\mathbf{S}}\mathbf{X}\,/\!\sim$$

Lemma 6

with

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

$\mathbb{L}$

Then, directly by definition,

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

$$\left(\mathbb{D}\mathbf{f}_X(\triangle\mathbf{P}), \mathbb{D}\mathbf{f}_S(\triangle\mathbf{P})\right) = -\mathbb{L}^{-1}\left(\triangle\mathbb{P}(\mathbf{X},\mathbf{S}), 0\right). \tag{20}$$

Defining

$$\|\triangle\mathbf{P}\| := \left\|[\mathbf{E}_0, \mathbf{E}_1, \ldots, \mathbf{E}_\ell]\right\|_F, \tag{21}$$

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

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

where the addition of pairs is unders(S)-1.23753.75977Td [(O)4.11.75977Td (y)-0.2.97011Tf -236.431(p)-38

Lemma 10 *Let* $(\mathbf{Y},\mathbf{S})$, $\mathbf{Y}\in\mathbb{C}^{\ell n\times k}$, $\mathbf{S}\in\mathbb{C}^{k\times k}$ *be a simple invariant pair of* $\mathbf{L}(\boldsymbol{\lambda})\in\mathbb{L}_1(\mathbf{P})$ *and let* $\mathbf{Y}$ *be partitioned as* $\mathbf{Y}=\left[\mathbf{Y}_\ell^{\mathsf{H}}\ \ \ldots\ \ \mathbf{Y}_1^{\mathsf{H}}\right]^{\mathsf{H}}$ *with* $\mathbf{Y}_j\in\mathbb{C}^{n\times k}$, $\mathbf{j}=1,\ldots,\ell$. *Then, for any* $\mathbf{j}\in[2,\ell]$, $(\mathbf{Y}_j,\mathbf{S})$ *is a simple invariant pair of* $\mathbf{P}(\boldsymbol{\lambda})$ *if and only if* $\mathbf{S}$ *is nonsingular.*

Proof. Theorem 9 implies that $(\mathbf{Y}_1,\mathbf{S})$ is a simple invariant pair and $\mathbf{Y}_j=\mathbf{Y}_1\mathbf{S}^{j-1}$. We obtain

$$\mathbb{P}(\mathbf{Y}_j,\mathbf{S})=\mathbf{A}_\ell\mathbf{Y}_j\mathbf{S}^\ell+\cdots+\mathbf{A}_1\mathbf{Y}_j\mathbf{S}+\mathbf{A}_0\mathbf{Y}_j=\mathbb{P}(\mathbf{Y}_1,\mathbf{S})\mathbf{S}^{j-1}=0.$$

If $\mathbf{S}$ is nonsingular then this relation implies that $(\mathbf{Y}_j,\mathbf{S})$ is an invariant pair. Moreover,

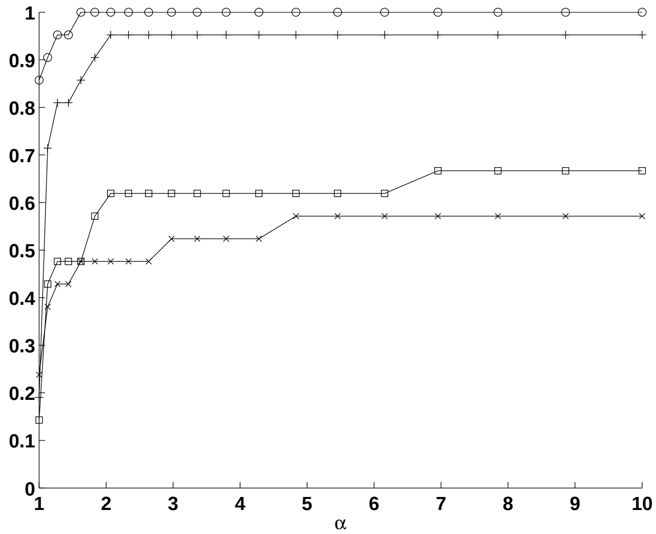
$$\operatorname{rank}(\mathbf{V}_\ell(\mathbf{Y}_j,\mathbf{S}))=\operatorname{rank}\left(\mathbf{V}_\ell(\mathbf{Y}_1,\mathbf{S})\mathbf{S}^{j-1}\right)\tag{28}$$

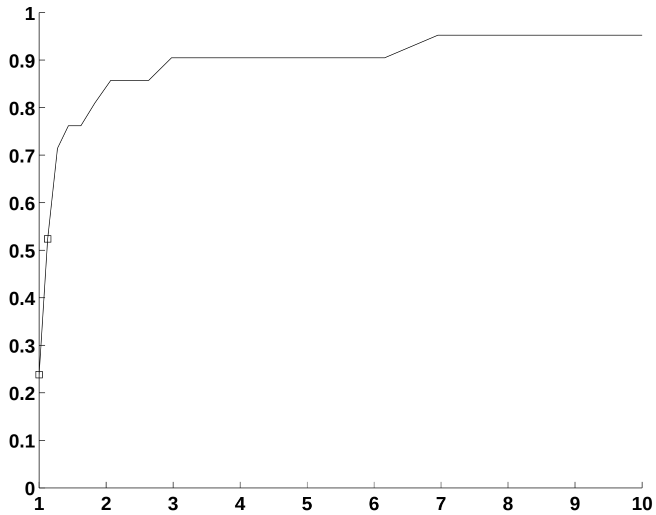
shows that $(\mathbf{Y}_j,\mathbf{S})$ is minimal and therefore a simple invariant pair. If $\mathbf{S}$ is singular then, by (28), $(\mathbf{Y}_j,\mathbf{S})$ is not minimal and is therefore not a simple invariant pair. □

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

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

$$\mathbf{R}(\hat{\mathbf{X}},\tilde{\mathbf{S}}) = \mathbf{y}$$


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9 10
α


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10

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 $(\mathbf{X}_0, \mathbf{S}_0)$ to a simple invariant pair $(\mathbf{X}, \mathbf{S}) \in \mathbb{C}^{n\times k} \times \mathbb{C}^{k\times k}$ our aim is to compute a correction that brings $(\mathbf{X}_0, \mathbf{S}_0)$ closer to $(\mathbf{X}, \mathbf{S})$. By Theorem 7, $(\mathbf{X}, \mathbf{S})$ is a regular value of the nonlinear matrix equations

$$\mathbb{P}(\mathbf{X}, \mathbf{S}) = 0, \quad \mathbb{V}(\mathbf{X}, \mathbf{S}) = 0, \tag{30}$$

where $\mathbb{P}(\mathbf{X}, \mathbf{S}) = \mathbf{X}\mathbf{A}_0 + \mathbf{X}\mathbf{A}_1\mathbf{S} + \cdots + \mathbf{X}\mathbf{A}_\ell\mathbf{S}^\ell$ and $\mathbb{V}(\mathbf{X}, \mathbf{S}) = \mathbf{W}^{\mathsf{H}}\mathbf{V}_m(\mathbf{X}, \mathbf{S}) - \mathbf{I}$ for some normalization matrix $\mathbf{W}^{\mathsf{H}} = [\mathbf{W}^{\mathsf{H}}_{m-1}, \ldots, \mathbf{W}^{\mathsf{H}}_0] \in \mathbb{C}^{k\times mn}$. Newton's method applied to (30) with starting value $(\mathbf{X}_0, \mathbf{S}_0)$ takes the form

$$(\mathbf{X}_{p+1}, \mathbf{S}_{p+1}) = (\mathbf{X}_p, \mathbf{S}_p) - \mathbb{L}_p^{-1}\big(\mathbb{P}(\mathbf{X}_p, \mathbf{S}_p), \mathbb{V}(\mathbf{X}_p, \mathbf{S}_p)\big), \tag{31}$$

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

$$\mathbb{L}_p(\triangle\mathbf{X}, \triangle\mathbf{S}) = \Big(\mathbb{P}(\triangle\mathbf{X}, \mathbf{S}_p) + \sum_{j=1}^{\ell} \mathbf{A}_j\mathbf{X}_p\,\mathbb{D}\mathbf{S}_p^j(\triangle\mathbf{S}),\ \sum_{j=0}^{m-1}\mathbf{W}_j^{\mathsf{H}}\big(\triangle\mathbf{X}\,\mathbf{S}_p^j + \mathbf{X}\,\mathbb{D}\mathbf{S}_p^j(\triangle\mathbf{S})\big)\Big),$$

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 $(\mathbf{X}_0, \mathbf{S}_0)$ is sufficiently close to $(\mathbf{X}, \mathbf{S})$.

In our implementation of (31) we keep the columns of $\mathbf{V}_m(\mathbf{X}_p, \mathbf{S}_p)$ orthonormal and adapt $\mathbf{W}$ correspondingly in the course of the iteration. For this purpose, we compute a (compact) QR decomposition

$$\mathbf{V}_m(\mathbf{X}_p, \mathbf{S}_p) = \mathbf{Q}\mathbf{R}$$

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

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_0\mathbf{R}\mathbf{S}_p^{m-1}\mathbf{R}^{-1} \\ \vdots \\ \mathbf{Q}_0\mathbf{R}\mathbf{S}_p\mathbf{R}^{-1} \\ \mathbf{Q}_0 \end{bmatrix}.$$

for $\mathbf{Q}_0 \in \mathbb{C}^{n\times k}$. Hence the replacement $(\mathbf{X}_p, \mathbf{S}_p) \leftarrow (\mathbf{Q}_0, \mathbf{R}\mathbf{S}_p\mathbf{R}^{-1})$ results in orthonormal $\mathbf{V}_m(\mathbf{X}_p, \mathbf{S}_p)$. Moreover, by (A) 1970 11 12 0 7 9 5 4 (t) 3.8 8 5 0 9 2 (1) 0.5 0 4 7 1 8 (e) 6.2 0 2 6 8 (O) (X) -1 1 5 2 4 2 4 0 1 3 5 5 9 0 7 9 0 9 (1)

Letting $\mathbf{r}_2$ and $\mathbf{q}_2$ denote the first columns of $\widetilde{\mathrm{Res}}_2$ and $\widetilde{\mathrm{Ort}}_2$, respectively, this shows that the second columns $\triangle\mathbf{x}_2, \triangle\mathbf{s}_2$ of $\triangle\mathbf{X}, \triangle\mathbf{S}$ satisfy the linear system

$$\begin{bmatrix} \mathbf{P}(\mathbf{s}_{22}) & \sum_{j=1}^{\ell} \mathbf{A}_j \mathbf{X} [\mathbb{D}\mathbf{S}^j]_{22} \\ \sum_{j=0}^{m-1} \mathbf{s}_{22}^j \mathbf{W}_j^{\mathsf{H}} & \sum_{j=1}^{m-1} \mathbf{W}_j^{\mathsf{H}} \mathbf{X} [\mathbb{D}\mathbf{S}^j]_{22} \end{bmatrix} \begin{bmatrix} \triangle\mathbf{x}_2 \\ \triangle\mathbf{s}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_2 \\ \mathbf{q}_2 \end{bmatrix}, \tag{37}$$

where $\mathbf{s}_{22}$ denotes the first diagonal element of $\mathbf{S}_{22}$ and $[\mathbb{D}\mathbf{S}^j]_{22}$ satisfies the recursion (34) with $\mathbf{s}_{11}$ replaced by $\mathbf{s}_{22}$.

The described process can be continued in an analogous manner to compute all columns of $\triangle\mathbf{X}$ and $\triangle\mathbf{S}$. The cost of the overall algorithm is dominated by the solution of $\mathbf{k}$ linear systems of the form (33) and (37). Since each of these systems has order with unitary matrices $\mathbf{Q}, \mathbf{Z} \in \mathbb{C}^{\ell n \times \ell n}$ and upper triangular matrices $\mathbf{T_A}, \mathbf{T_B} \in \mathbb{C}^{\ell n \times \ell n}$. Note that if the initial approximation $(\mathbf{X}_0, \mathbf{S}_0$

implementation [31] 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

We perturb

[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*

[23] L. Ho nung, 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. Technical report 2009-05, Seminar for applied mathematics, ETH Zurich, February 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] Karl Meerbergen. Locking and restarting quadratic eigenvalue solvers. *SIAM J. Sci. Comput.*, 22(5):1814–1839 (electronic), 2000.

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

[32] 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.

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

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

[35] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic5(n)49(c-2.66764(i)2.77954(c5(n)

and hence

$$
\begin{aligned}
(\mathcal{A} + \boldsymbol{\lambda}\mathcal{B})\mathbf{y}_j &= \sum_{i=1}^{j} \frac{1}{(\mathbf{i}-1)!}(\mathcal{A} + \boldsymbol{\lambda}\mathcal{B})\mathbf{V}_\ell^{(i-1)}(\mathbf{x}_{j-i+1}, \boldsymbol{\lambda}) \\
&= \sum_{i=1}^{j} \left(\mathbf{v} \otimes \frac{1}{(\mathbf{i}-1)!}\mathbf{P}^{(i-1)}(\boldsymbol{\lambda})\mathbf{x} - \frac{(\mathbf{i}-1)}{(\mathbf{i}-1)!}\mathbf{B}\mathbf{V}_\ell^{(i-2)}(\mathbf{x}, \boldsymbol{\lambda}) \right) \\
&= -\sum^{j} \Big)
\end{aligned}
$$