Numer. Math. (2010) 116:95–122
DOI 10.1007/s00211-010-0297-4

Numerische Mathematik

Structured backward error for palindromic polynomial eigenvalue problems

Ren-Cang Li · Wen-Wei Lin · Chern-Shuh Wang

Received: 8 September 2008 / Revised: 5 February 2010 / Published online: 24 March 2010

Abstract A detailed structured backward error analysis for four kinds of palindromic polynomial eigenvalue problems (PPEP)

$$\left(\sum_{\ell=0}^{d} A_\ell \lambda^\ell\right) x = 0, \quad A_{d-\ell} = \varepsilon A_\ell^\star \text{ for } \ell = 0, 1, \ldots, \lfloor d/2 \rfloor,$$

where $\star$ is one of the two actions: transpose and conjugate transpose, and $\varepsilon \in \{\pm 1\}$. Each of them has its application background with the case $\star$ taking transpose and $\varepsilon = 1$ attracting a great deal of attention lately because of its application in the fast train modeling. Computable formulas and bounds for the structured backward errors are obtained. The analysis reveals distinctive features of PPEP from general polynomial eigenvalue problems (PEP) investigated by Tisseur (Linear Algebra Appl 309:339–361, 2000) and by Liu and Wang (Appl Math Comput 165:405–417, 2005).

Mathematics Subject Classification (2000) 65F15 · 15A18

R.-C. Li (✉)
Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019, USA
e-mail: rcli@uta.edu

W.-W. Lin
Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan
e-mail: wwlin@math.nctu.tw; wwlin@math.edu.tw

C.-S. Wang
Department of Mathematics, National Cheng Kung University, Tainan 701, Taiwan
e-mail: cswang@math.ncku.tw; cswang@math.edu.tw

1 Introduction

In vibration analysis of fast trains arises the palindromic quadratic eigenvalue problem (PQEP) [4,12,14]: find a scalar $\lambda \in \mathbb{C}$ and an n-dimensional vector $x \in \mathbb{C}^n \setminus \{0\}$ such that

$$Q(\lambda)x \equiv (A_0^\top \lambda^2 + A_1\lambda + A_0)x = 0, \tag{1.1}$$

where A_0 and A_1 are $n \times n$ (real or complex) matrices with $A_1^\top = A_1$, and the superscript "$\cdot^\top$" takes matrix transpose. The scalar λ and the nonzero vector x in (1.1) are an *eigenvalue* and its *associated eigenvector* of $Q(\lambda)$. The pair $\{\lambda, x\}$ is called an *eigenpair* of $Q(\lambda)$. Presently, PQEP is solved by certain kinds of structure-preserving algorithms, such as Jacobi-type method [9,15], QR-like algorithm [16], Patel's algorithm [11] and URV-like algorithm [17], as well as the structure-preserving doubling algorithm [4]. Experience shows that PQEP from the fast train application is notoriously difficult to solve accurately, partly because of its widely varying magnitudes in eigenvalues. How to solve it accurately is still an active research topic.

PQEP is a particular case of the so-called *Polynomial Eigenvalue Problem* (PEP) of degree d:

$$\left(\sum_{\ell=0}^{d} A_\ell \lambda^\ell\right) x = 0, \tag{1.2}$$

where A_ℓ, $\ell = 0, 1, \ldots, d$ are $n \times n$ complex matrices. When $d = 2$, PEP (1.2) is called the quadratic eigenvalue problem (QEP) [20]. For a PEP in its generality there is no relation among coefficient matrices, unlike PQEP. Naturally one defines the palindromic polynomial eigenvalue problem (PPEP) to be a PEP (1.2) with $A_{d-\ell} = A_\ell^\top$, $\ell = 0, 1, \ldots, \lfloor d/2 \rfloor$, namely,

$$P(\lambda)x \equiv \left(\sum_{\ell=0}^{d} A_\ell \lambda^\ell\right) x = 0, \quad A_{d-\ell} = A_\ell^\top, \quad \text{for } \ell = 0, 1, \ldots, \lfloor d/2 \rfloor, \tag{1.3}$$

where $\lfloor d/2 \rfloor$ is the largest integer that is no bigger than $d/2$.

Let $\{\mu, z\}$ be a computed eigenpair of (1.2) or (1.3). Ideally, we would like to have $\left(\sum_{\ell=0}^{d} A_\ell \mu^\ell\right) z = 0$. But practically we have $\left(\sum_{\ell=0}^{d} A_\ell \mu^\ell\right) z = -r$ with residual $r \neq 0$ but usually tiny. Backward error analysis asks if the computed eigenpair $\{\mu, z\}$ is an exact eigenpair of a *nearby* PEP, namely,

$$\left(\sum_{\ell=0}^{d} (A_\ell + \rho_\ell \Delta A_\ell)\mu^\ell\right) z = 0$$

with *backward perturbation matrices* ΔA_ℓ ($\ell = 0, 1, \ldots, d$) hopefully tiny in magnitude, where ρ_ℓ ($\ell = 0, 1, \ldots, d$) are *scaling parameters*, usually taken to be some norms of A_ℓ, respectively. For a general PEP, no constraints on and/or among ΔA_ℓ, $\ell = 0, 1, \ldots, d$ are imposed, except that ΔA_ℓ, $\ell = 0, 1, \ldots, d$ should have as tiny magnitude as possible. But when a PEP has worthy structures like PPEP, naturally we would ask if the computed eigenpair $\{\mu, z\}$ is an exact eigenpair of a *nearby* PPEP. This is the so-called *structured backward perturbation analysis* associated with the approximate eigenpair for PPEP.

Tisseur [18] thoroughly developed a backward perturbation analysis for PEP in its generality, where no structure upon ΔA_ℓ is imposed because there is not any on A_ℓ assumed. When A_ℓ, $\ell = 0, 1, \ldots, d$ are Hermitian, naturally one would like to enforcing ΔA_ℓ Hermitian, too. This is the *structured backward perturbation analysis* for the case. In the same article, Tisseur obtained a result when μ is real. When μ is complex, the structured backward perturbation analysis is completed by Liu and Wang [13] who also investigated the corresponding question for the generalized and multi-parameter eigenvalue problem. Two other related works are Higham and Higham [6] and Hochstenbach and Plestenjak [10]. None of the results from these papers apply to PPEP's structured backward perturbation analysis, however. Recently Bora [1] investigated under what circumstances the structured backward errors are equal to the unstructured backward errors for PPEP, but did not actually obtain computable formulas.

In this paper, we shall consider PPEP in a broader sense, with (1.3) being one of the four different kinds. To present these PPEP compactly, we let the superscript "$\cdot^\star$" be either "$\cdot^\top$" or "$\cdot^{\mathrm{H}}$" which takes complex conjugate and transpose and let $\varepsilon \in \{\pm 1\}$. Collectively by a PPEP, we mean a polynomial eigenvalue problem of the following form

$$\left(\sum_{\ell=0}^{d} A_\ell \lambda^\ell\right) x = 0, \quad A_{d-\ell} = \varepsilon A_\ell^\star \quad \text{for } \ell = 0, 1, \ldots, \lfloor d/2 \rfloor. \tag{1.4}$$

We have mentioned the fast train application [4,12,14] which yielded a problem of this form with $d = 2$, $\star = \top$, and $\varepsilon = 1$. It was first raised in the study of the vibration in the structural analysis for fast trains [8,9], and then of the behavior of periodic surface acoustic wave (SAW) filters [22]. PQEP with $\star = \mathrm{H}$ was raised in the computation of the Crawford number by the bisection and level set methods in [7]. A typical even degree PPEP of (1.4) can be derived from solving linear quadratic discrete-time optimal control problem subject to the higher order discrete system [3,21].

Let $\{\mu, z\}$ be a computed eigenpair and let $\|\cdot\|$ be either the spectral norm[1] $\|\cdot\|_2$ or the Frobenius norm $\|\cdot\|_{\mathrm{F}}$. We are interested in knowing the structured backward error

$$\begin{array}{ll} \Delta = \min \sqrt{\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta A_\ell\|^2}, & \\ \text{subject to} & \left(\sum_{\ell=0}^{d} (A_\ell + \rho_\ell \Delta A_\ell)\mu^\ell\right) z = 0, \\ & \text{scaling parameters } \rho_{d-\ell} = \rho_\ell \geq 0, \\ & A_{d-\ell} + \rho_{d-\ell} \Delta A_{d-\ell} = \varepsilon (A_\ell + \rho_\ell \Delta A_\ell)^\star, \\ & \text{for } \ell = 0, 1, \ldots, \lfloor d/2 \rfloor. \end{array} \tag{1.5}$$

[1] Notation $\|\cdot\|_2$ is generic and on a vector it is its ℓ_2-norm.

To distinguish the norm used, later we will attach a subscript "2" or "F" to the structured backward error Δ in (1.5). In the usual PEP case, backward errors always exist. But as one might expect, forcing backward perturbations to be structured creates critical cases where backward errors fail to exist, i.e., there are no ΔA_ℓ, $\ell = 0, 1, \ldots, d$, satisfying the constraints in (1.5). In fact, as we shall see, the critical cases are when $|\mu| = 1$ if $\star = \mathrm{H}$ and when $\mu = \pm 1$ if $\star = \mathrm{T}$.

The rest of this paper is organized as follows. Section 2 significantly simplifies (1.5) to a problem of solving a 2-by-2 matrix equation, paving the way for us to derive the backward errors for $\star = \mathrm{H}$ in Sect. 3 and for $\star = \mathrm{T}$ in Sect. 4. It turns out the case when $\star = \mathrm{H}$ is considerably more complicated and interesting than the case when $\star = \mathrm{T}$. Numerical examples are given in Sect. 5 to illustrate our results, and finally we present a few concluding remarks in Sect. 6.

Notation Throughout this paper, $\mathbb{C}^{n\times m}$ is the set of all $n \times m$ complex matrices, $\mathbb{C}^n = \mathbb{C}^{n\times 1}$, and $\mathbb{C} = \mathbb{C}^1$. Similarly define $\mathbb{R}^{n\times m}$, $\mathbb{R}^n$, and $\mathbb{R}$ except replacing the word *complex* by *real*. I_n (or simply I if its dimension is clear from the context) is the $n \times n$ identity matrix. $\bar{x}$ is the complex conjugate of a scalar or the entry-wise complex conjugate of a vector or matrix. $\Re(\alpha)$ and $\Im(\alpha)$ are the real and imaginary part of α, respectively, and ι is the imaginary unit. We shall also adopt MATLAB-like convention to access the entries of vectors and matrices. $i : j$ is the set of integers from i to j inclusive and $i : i = \{i\}$. For a matrix X, $X_{(i,j)}$ is X's (i, j)th entry, and $X_{(k:\ell,i:j)}$ is X's submatrices consisting of intersections of row k to row ℓ and column i to column j. $X^\dagger$ is X's Moore–Penrose inverse.

2 Simplify the problem

The constraints in (1.5) require $\rho_{d-\ell} = \rho_\ell$ and $\Delta A_{d-\ell} = \varepsilon \Delta A_\ell^\star$, for $\ell = 0, 1, \ldots, \lfloor d/2 \rfloor$, and

$$\left(\sum_{\ell=0}^{d} \rho_\ell \Delta A_\ell \mu^\ell\right) z = r \stackrel{\text{def}}{=} -\left(\sum_{\ell=0}^{d} A_\ell \mu^\ell\right) z, \tag{2.1}$$

where $\Delta A_\ell \in \mathbb{C}^{n\times n}$ ($\ell = 0, 1, \ldots, d$). We will seek if (2.1) has a solution $\{\Delta A_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ and if it does, we'll seek ΔA_ℓ such that

$$\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta A_\ell\|^2 \text{ is minimized,} \tag{2.2}$$

where $\|\cdot\|$ is either $\|\cdot\|_2$ or $\|\cdot\|_{\mathrm{F}}$. The two trivial cases (i) when all $\rho_\ell = 0$, and (ii) when $r = 0$ will be excluded from our consideration.

The key to solve (2.1) and (2.2) is a reduction technique that has been used in, e.g., [2,5,13,19]. The technique allows us to transform (2.1) to essentially the case when $n = 2$ by focusing on backward errors in a 2-dimensional subspace spanned by $\{z, r\}$ or by $\{z, \bar{r}\}$.

Assume $n \geq 2$. Later we shall comment on which part of our analysis will cover the case when $n = 1$, i.e., the scalar polynomial case. See Remarks 3.5 and 4.1.

Let $Q \in \mathbb{C}^{n\times n}$ be a unitary matrix, i.e., $QQ^{\mathrm{H}} = I_n$, such that

$$Q^{\mathrm{H}}(z\ \widehat{r}) = \begin{pmatrix} \alpha & \gamma \\ 0 & \beta \\ \hline 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \tag{2.3}$$

where

$$\widehat{r} = r \quad \text{if } \star = \mathrm{H}, \quad \text{and} \quad \widehat{r} = \bar{r} \quad \text{if } \star = \top. \tag{2.4}$$

Such Q always exists, and necessarily (2.3) implies

$$|\alpha| = \|z\|_2,\ \gamma = \frac{z^{\mathrm{H}}\widehat{r}}{\bar{\alpha}}, \quad |\beta| = \left\| \widehat{r} - \frac{z^{\mathrm{H}}\widehat{r}}{\|z\|_2^2} z \right\|_2 = \frac{\sqrt{\|\widehat{r}\|_2^2\|z\|_2^2 - |z^{\mathrm{H}}\widehat{r}|^2}}{\|z\|_2}. \tag{2.5}$$

It follows from (2.1) that

$$Q^{\star}\left(\sum_{\ell=0}^{d} \rho_\ell \Delta A_\ell \mu^\ell\right) QQ^{\mathrm{H}}z = Q^{\star}r, \tag{2.6}$$

or equivalently

$$\left(\sum_{\ell=0}^{d} \rho_\ell \Delta B_\ell \mu^\ell\right) y = w,\ \rho_{d-\ell} = \rho_\ell \quad \text{and} \quad \Delta B_{d-\ell} = \varepsilon \Delta B_\ell^{\star}, \quad \text{for } \ell = 0, 1, \ldots, \lfloor d/2 \rfloor, \tag{2.7}$$

where

$$\Delta B_\ell = Q^{\star}(\Delta A_\ell)Q, \quad y = Q^{\mathrm{H}}z = \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad w = Q^{\star}r = \begin{pmatrix} \widehat{\gamma} \\ \widehat{\beta} \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$\widehat{\beta} = \beta \quad \text{and} \quad \widehat{\gamma} = \gamma \quad \text{if } \star = \mathrm{H}, \quad \text{and} \quad \widehat{\beta} = \bar{\beta} \quad \text{and} \quad \widehat{\gamma} = \bar{\gamma} \quad \text{if } \star = \top.$$

Since Q is unitary, (2.1) and (2.7) have the same solvability property, i.e., if one is solvable, so is the other, and moreover

$$\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta B_\ell\|^2 = \sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta A_\ell\|^2.$$

Thus the optimal solution $\{\Delta B_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ to (2.7) in the sense that

$$\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta B_\ell\|^2 \text{ is minimized} \tag{2.8}$$

gives rise to one solution $\{\Delta A_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ to (2.1) in the sense of (2.2), and vice versa. Set

$$\delta_1 = \widehat{\gamma}/\alpha, \quad \delta_2 = \widehat{\beta}/\alpha. \tag{2.9}$$

and

$$\Delta_p = \min\left\{ \sqrt{\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta B_\ell\|_p^2} \ :\ (2.7) \text{ satisfied} \right\} \quad \text{for } p = 2, \text{F}. \tag{2.10}$$

$\Delta_p = \infty$ means that (2.7) is not solvable. It follows from (2.5) that

$$|\delta_1| = \frac{|z^{\mathrm{H}}\widehat{r}|}{\|z\|_2^2}, \quad |\delta_2| = \frac{\sqrt{\|\widehat{r}\|_2^2\|z\|_2^2 - |z^{\mathrm{H}}\widehat{r}|^2}}{\|z\|_2^2}, \quad \sqrt{|\delta_1|^2 + |\delta_2|^2} = \frac{\|r\|_2}{\|z\|_2}. \tag{2.11}$$

In particular

$$\delta_1 = \frac{\gamma}{\alpha} = \frac{z^{\mathrm{H}} r}{\bar{\alpha}\alpha} = \frac{z^{\mathrm{H}} r}{\|z\|_2^2} \quad \text{for } \star = \text{H}. \tag{2.12}$$

Theorem 2.1 *Suppose that (2.7) has a solution (so does (2.1)). The optimal solution $\{\Delta B_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ to (2.7) in the sense of (2.8) for the Frobenius norm satisfies*

$$(\Delta B_\ell)_{(i,j)} = 0 \ \text{ for } i,\ j > 2 \ \text{ and } \ \text{for } \ell = 0, 1, \ldots, \lfloor d/2 \rfloor. \tag{2.13}$$

There is an optimal solution $\{\Delta B_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ in the sense of (2.8) for the spectral norm that satisfies (2.13). Finally

$$\Delta_2 \le \Delta_{\mathrm{F}} \le \sqrt{2}\,\Delta_2. \tag{2.14}$$

Proof No proof is needed for those ΔB_ℓ with $\rho_\ell = 0$. Suppose all ρ_ℓ are positive.

Equating the corresponding entries at the both sides of (2.7) leads to n linear equations in the entries of ΔB_ℓ, $\ell = 0, 1, \ldots, \lfloor d/2 \rfloor$. The last $n-2$ equations are homogenous and contain only those entries

$$(\Delta B_\ell)_{(i,j)} = 0 \quad \text{for } i,\ j > 2 \quad \text{and} \quad \text{for } \ell = 0, 1, \ldots, \lfloor d/2 \rfloor$$

and at the same time they do not appear in the first 2 equations. Therefore the optimal solution $\{\Delta B_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ in the sense of (2.8) for the Frobenius norm must satisfy (2.13).

For any 2-by-2 block matrix

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix},$$

we have $\|X\|_2 \geq \|X_{11}\|_2$. Therefore if $\{\Delta B_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ is an optimal solution to (2.7) for the spectral norm, resetting $(\Delta B_\ell)_{(i,j)} = 0$ for $i, j > 2$ and $\ell = 0, 1, \ldots, \lfloor d/2 \rfloor$ and leaving $(\Delta B_\ell)_{(i,j)}$ for $1 \leq i, j \leq 2$ and $\ell = 0, 1, \ldots, \lfloor d/2 \rfloor$ untouched yield a solution to (2.7) and at the same time do not make $\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta B_\ell\|_2^2$ bigger. Since before resetting, $\{\Delta B_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ is optimal, $\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta B_\ell\|_2^2$ will not change before and after the resetting. Thus the resulting $\{\Delta B_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ is an optimal solution satisfying (2.13) for the spectral norm.

Let $\{\Delta B_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ be the optimal solution in the Frobenius norm. Then

$$\Delta_2 \leq \sqrt{\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta B_\ell\|_2^2} \leq \sqrt{\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta B_\ell\|_{\mathrm{F}}^2} = \Delta_{\mathrm{F}}.$$

On the other hand, let $\{\Delta B_\ell, \ell = 0, 1, \ldots, \lfloor d/2 \rfloor\}$ be an optimal solution satisfying (2.13) in the spectral norm. Then $\operatorname{rank}(\Delta B_\ell) \leq 2$ for $\ell = 0, 1, \ldots, \lfloor d/2 \rfloor$. Thus

$$\Delta_{\mathrm{F}} \leq \sqrt{\sum_{\ell=0}^{\lfloor d/2 \rfloor} \|\Delta B_\ell\|_{\mathrm{F}}^2} \leq \sqrt{\sum_{\ell=0}^{\lfloor d/2 \rfloor} 2\|\Delta B_\ell\|_2^2} = \sqrt{2}\,\Delta_2,$$

as expected. □

Theorem 2.1 tells us that it suffices to consider these two equations derived from the first two entries at both sides of $\left(\sum_{\ell=0}^{d} \rho_\ell \Delta B_\ell \mu^\ell\right) y = w$ in (2.7). In the two equations, only the entries $(\Delta B_\ell)_{(i,j)}$ for $1 \leq i, j \leq 2$ and $\ell = 0, 1, \ldots, \lfloor d/2 \rfloor$ appear. We shall do so in the next two sections for each different combination of the superscript $\star \in \{\top, \mathrm{H}\}$ and $\varepsilon \in \{\pm 1\}$. For this purpose, let

$$(\Delta B_\ell)_{(1:2,1:2)} = \begin{pmatrix} b_{11}^{(\ell)} & b_{12}^{(\ell)} \\ b_{21}^{(\ell)} & b_{22}^{(\ell)} \end{pmatrix}. \tag{2.15}$$

Constraints among $b_{ij}^{(\ell)}$ because of $\Delta B_{d-\ell} = \varepsilon \Delta B_\ell^\star$, for $\ell = 0, 1, \ldots, \lfloor d/2 \rfloor$ will be imposed at the time we consider each case.

Recall in (2.7) $\rho_{d-\ell} = \rho_\ell$ and $\Delta B_{d-\ell} = \varepsilon \Delta B_\ell^\star$ for $\ell = 0, 1, \dots, \lfloor d/2 \rfloor$. For even d, we have

$$\begin{aligned}
\sum_{\ell=0}^{d} \rho_\ell \Delta B_\ell \mu^\ell &= \sum_{\ell=0}^{d/2-1} \rho_\ell \Delta B_\ell \mu^\ell + \rho_{d/2} \Delta B_{d/2} \mu^{d/2} + \sum_{\ell=d/2+1}^{d} \rho_\ell \Delta B_\ell \mu^\ell \\
&= \sum_{\ell=0}^{d/2-1} \rho_\ell \Delta B_\ell \mu^\ell + \rho_{d/2} \Delta B_{d/2} \mu^{d/2} + \sum_{j=d/2-1}^{0} \rho_{d-j} \Delta B_{d-j} \mu^{d-j} \\
&= \sum_{\ell=0}^{d/2-1} \rho_\ell \Delta B_\ell \mu^\ell + \rho_{d/2} \Delta B_{d/2} \mu^{d/2} + \sum_{j=0}^{d/2-1} \rho_j \varepsilon \Delta B_j^\star \mu^{d-j} \\
&= \sum_{\ell=0}^{d/2-1} \rho_\ell \left(\Delta B_\ell + \varepsilon \Delta B_\ell^\star \mu^{d-2\ell} \right) \mu^\ell + \rho_{d/2} \Delta B_{d/2} \mu^{d/2} \qquad (2.16)
\end{aligned}$$

and $\Delta B_{d/2} = \varepsilon \Delta B_{d/2}^\star$. For odd d, we have

$$\begin{aligned}
\sum_{\ell=0}^{d} \rho_\ell \Delta B_\ell \mu^\ell &= \sum_{\ell=0}^{(d-1)/2} \rho_\ell \Delta B_\ell \mu^\ell + \sum_{\ell=(d-1)/2+1}^{d} \rho_\ell \Delta B_\ell \mu^\ell \\
&= \sum_{\ell=0}^{(d-1)/2} \rho_\ell \Delta B_\ell \mu^\ell + \sum_{j=(d-1)/2}^{0} \rho_{d-j} \Delta B_{d-j} \mu^{d-j} \\
&= \sum_{\ell=0}^{(d-1)/2} \rho_\ell \Delta B_\ell \mu^\ell + \sum_{j=0}^{(d-1)/2} \rho_j \varepsilon \Delta B_j^\star \mu^{d-j} \\
&= \sum_{\ell=0}^{(d-1)/2} \rho_\ell \left(\Delta B_\ell + \varepsilon \Delta B_\ell^\star \mu^{d-2\ell} \right) \mu^\ell. \qquad (2.17)
\end{aligned}$$

Our detailed analysis in the next two sections is for the Frobenius norm only. Bounds on Δ_2 for the spectral norm can then be deduced straightforwardly by using (2.14). So in the next two sections, $\|\cdot\|$ in (2.8) is always the Frobenius norm. Recall the assumption $r \neq 0$ which implies

$$|\delta_1| + |\delta_2| > 0. \qquad (2.18)$$

For the ease of analysis, we assume also

$$\text{all } \rho_\ell > 0. \qquad (2.19)$$

This will remove many special cases caused by one or more ρ_ℓ possibly being zeros. Note for a particular ℓ, if $\rho_\ell = 0$, then must $\Delta A_\ell = 0$ by (2.2). Thus the assumption (2.19) limits the applicability of our analysis to cases where no backward perturbations to some of the A_ℓ's are preferred. But we argue that such cases can be practically dealt with satisfactorily by giving sufficiently tiny positive numbers to the corresponding ρ_ℓ's. Indeed, our later formulas for the optimal ΔA_ℓ (see Lemmas 3.1–3.4 and 4.1–4.4) is linear in and proportional to ρ_ℓ which means for any particular ℓ, letting ρ_ℓ go to 0 implies the corresponding optimal ΔA_ℓ going to 0 as well.

The following lemma will be needed in the later sections.

Lemma 2.1 *(i) For $\tau = |\tau|e^{\iota\vartheta} \in \mathbb{C}$, $\vartheta \in \mathbb{R}$, and $\eta \in \mathbb{C}$, we have*

$$\begin{pmatrix} \Re(\tau\,\eta) \\ \Im(\tau\,\eta) \end{pmatrix} = |\tau| \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix} \begin{pmatrix} \Re(\eta) \\ \Im(\eta) \end{pmatrix},$$
$$\begin{pmatrix} \Re(\tau\,\bar\eta) \\ \Im(\tau\,\bar\eta) \end{pmatrix} = |\tau| \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix} \begin{pmatrix} \Re(\eta) \\ \Im(\eta) \end{pmatrix}.$$

(ii) For $\xi \in \mathbb{R}$ and $\vartheta \in \mathbb{R}$, we have

$$\left[I_2 + \xi \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix} \right]^2 = (1+\xi^2) I_2 + 2\xi \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix},$$
$$\begin{pmatrix} \cos\vartheta \\ \sin\vartheta \end{pmatrix} \begin{pmatrix} \cos\vartheta \\ \sin\vartheta \end{pmatrix}^\top = \frac{1}{2} \left[I + \begin{pmatrix} \cos 2\vartheta & \sin 2\vartheta \\ \sin 2\vartheta & -\cos 2\vartheta \end{pmatrix} \right],$$
$$\begin{pmatrix} -\sin\vartheta \\ \cos\vartheta \end{pmatrix} \begin{pmatrix} -\sin\vartheta \\ \cos\vartheta \end{pmatrix}^\top = \frac{1}{2} \left[I - \begin{pmatrix} \cos 2\vartheta & \sin 2\vartheta \\ \sin 2\vartheta & -\cos 2\vartheta \end{pmatrix} \right].$$

3 Structured backward error for $\star = \mathrm{H}$

Theorem 3.1 is the main result of this section. We first define a few parameters in term of a given approximate eigenpair $\{\mu, z\}$ of PPEP (1.4): for even d,

$$\xi_{\text{even}} \overset{\text{def}}{=} \frac{\rho_{d/2}^2 |\mu|^d}{2} + \sum_{\ell=0}^{d/2-1} \rho_\ell^2 \left(|\mu|^{2\ell} + |\mu|^{2(d-\ell)} \right), \tag{3.1}$$

$$\zeta_{\text{even}} \overset{\text{def}}{=} |\mu|^d \left(\frac{\rho_{d/2}^2}{2} + 2 \sum_{\ell=0}^{d/2-1} \rho_\ell^2 \right), \tag{3.2}$$

$$\Phi_{\text{even}} \overset{\text{def}}{=} \xi_{\text{even}} + \zeta_{\text{even}} = \sum_{\ell=0}^{d/2-1} \rho_\ell^2 \left(|\mu|^\ell + |\mu|^{d-\ell} \right)^2 + \rho_{d/2}^2 |\mu|^d, \tag{3.3}$$

$$\phi_{\text{even}} \overset{\text{def}}{=} \xi_{\text{even}} - \zeta_{\text{even}} = \sum_{\ell=0}^{d/2-1} \rho_\ell^2 \left(|\mu|^\ell - |\mu|^{d-\ell} \right)^2, \tag{3.4}$$

$$\Psi_{\text{even}} \overset{\text{def}}{=} \sum_{\ell=0}^{d/2-1} \rho_\ell^2 \left(|\mu|^{2\ell} + |\mu|^{2(d-\ell)} \right) + \rho_{d/2}^2 |\mu|^d / 2, \tag{3.5}$$

and for odd d,

$$\xi_{\text{odd}} \stackrel{\text{def}}{=} \sum_{\ell=0}^{(d-1)/2} \rho_\ell^2 \left(|\mu|^{2\ell} + |\mu|^{2(d-\ell)}\right), \tag{3.6}$$

$$\zeta_{\text{odd}} \stackrel{\text{def}}{=} 2|\mu|^d \sum_{\ell=0}^{(d-1)/2} \rho_\ell^2, \tag{3.7}$$

$$\Phi_{\text{odd}} \stackrel{\text{def}}{=} \xi_{\text{odd}} + \zeta_{\text{odd}} = \sum_{\ell=0}^{(d-1)/2} \rho_\ell^2 \left(|\mu|^{\ell} + |\mu|^{d-\ell}\right)^2, \tag{3.8}$$

$$\phi_{\text{odd}} \stackrel{\text{def}}{=} \xi_{\text{odd}} - \zeta_{\text{odd}} = \sum_{\ell=0}^{(d-1)/2} \rho_\ell^2 \left(|\mu|^{\ell} - |\mu|^{d-\ell}\right)^2, \tag{3.9}$$

$$\Psi_{\text{odd}} \stackrel{\text{def}}{=} \sum_{\ell=0}^{(d-1)/2} \rho_\ell^2 \left(|\mu|^{2\ell} + |\mu|^{2(d-\ell)}\right). \tag{3.10}$$

Throughout this section, these assignments (3.1)–(3.10) are assumed.

Theorem 3.1 *Let $\{\mu, z\}$ be a given approximate eigenpair of PPEP (1.4). Suppose $\star = \mathrm{H}$ and $\varepsilon = \pm 1$ in (1.5), and δ_1 and δ_2 are as in (2.11) with $\widehat{r} = r$ which is defined in (2.1). Let*

$$\phi = \begin{cases} \phi_{\text{even}} & \textit{for even } d, \\ \phi_{\text{odd}}, & \textit{for odd } d, \end{cases} \quad \Phi = \begin{cases} \Phi_{\text{even}} & \textit{for even } d, \\ \Phi_{\text{odd}}, & \textit{for odd } d, \end{cases} \quad \Psi = \begin{cases} \Psi_{\text{even}} & \textit{for even } d, \\ \Psi_{\text{odd}}, & \textit{for odd } d. \end{cases}$$

For the structured backward error Δ_{F} defined in (1.5), we have

1. *If $|\mu| = 1$ and $z^{\mathrm{H}} r/(\sqrt{\varepsilon}\mu^{d/2}) \notin \mathbb{R}$, then $\Delta_{\mathrm{F}} = +\infty$.*
2. *If $|\mu| = 1$ and $z^{\mathrm{H}} r/(\sqrt{\varepsilon}\mu^{d/2}) \in \mathbb{R}$, then*

$$\Delta_{\mathrm{F}} = \sqrt{\frac{|\delta_1|^2}{\Phi} + \frac{|\delta_2|^2}{\Psi}}.$$

3. *If $|\mu| \neq 1$, then*

$$\Delta_{\mathrm{F}} \le \sqrt{\frac{|\delta_1|^2}{\phi} + \frac{|\delta_2|^2}{\Psi}}. \tag{3.11}$$

Remark 3.1 Inequality (3.11) in general cannot be improved because it can become an equality (see Remarks 3.2 and 3.4 below). Besides bounding Δ_{F} as in (3.11), a close formula for it for the case $|\mu| \neq 1$ at the price of a more complicated expression can also be gotten by combining (3.24), (3.42), (3.44), and (3.51) below. We omit the detail.

The proof of this theorem rests on solving (2.7). It is done by Lemmas 3.1–3.4 below. Recall (2.15), (2.16), (2.17), and Theorem 2.1, and keep in mind that $\star = \mathrm{H}$, $\varepsilon = \pm 1$, and $\|\cdot\| = \|\cdot\|_{\mathrm{F}}$ in (2.8).

For even d, we need to solve

$$\left[\sum_{\ell=0}^{d/2-1} \rho_\ell \left(\begin{pmatrix} b_{11}^{(\ell)} & b_{12}^{(\ell)} \\ b_{21}^{(\ell)} & b_{22}^{(\ell)} \end{pmatrix} + \varepsilon \begin{pmatrix} \bar{b}_{11}^{(\ell)} & \bar{b}_{21}^{(\ell)} \\ \bar{b}_{12}^{(\ell)} & \bar{b}_{22}^{(\ell)} \end{pmatrix} \mu^{d-2\ell} \right) \mu^\ell + \rho_{d/2} \begin{pmatrix} b_{11}^{(d/2)} & b_{12}^{(d/2)} \\ \varepsilon \bar{b}_{12}^{(d/2)} & b_{22}^{(d/2)} \end{pmatrix} \mu^{d/2} \right] \begin{pmatrix} \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma \\ \beta \end{pmatrix}, \tag{3.12}$$

where $\sqrt{\varepsilon} b_{11}^{(d/2)}$ and $\sqrt{\varepsilon} b_{22}^{(d/2)}$ are real (which means $b_{11}^{(d/2)}$ and $b_{22}^{(d/2)}$ are real for $\varepsilon = 1$ and pure imaginary for $\varepsilon = -1$). Componentwise, it gives

$$\sum_{\ell=0}^{d/2-1} \rho_\ell \left(b_{11}^{(\ell)} + \varepsilon \bar{b}_{11}^{(\ell)} \mu^{d-2\ell} \right) \mu^\ell + \rho_{d/2} b_{11}^{(d/2)} \mu^{d/2} = \delta_1, \tag{3.13}$$

$$\sum_{\ell=0}^{d/2-1} \rho_\ell \left(b_{21}^{(\ell)} + \varepsilon \bar{b}_{12}^{(\ell)} \mu^{d-2\ell} \right) \mu^\ell + \varepsilon \rho_{d/2} \bar{b}_{12}^{(d/2)} \mu^{d/2} = \delta_2. \tag{3.14}$$

For odd d, we need to solve

$$\left[\sum_{\ell=0}^{(d-1)/2} \rho_\ell \left(\begin{pmatrix} b_{11}^{(\ell)} & b_{12}^{(\ell)} \\ b_{21}^{(\ell)} & b_{22}^{(\ell)} \end{pmatrix} + \varepsilon \begin{pmatrix} \bar{b}_{11}^{(\ell)} & \bar{b}_{21}^{(\ell)} \\ \bar{b}_{12}^{(\ell)} & \bar{b}_{22}^{(\ell)} \end{pmatrix} \mu^{d-2\ell} \right) \mu^\ell \right] \begin{pmatrix} \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma \\ \beta \end{pmatrix}. \tag{3.15}$$

Componentwise, it gives

$$\sum_{\ell=0}^{(d-1)/2} \rho_\ell \left(b_{11}^{(\ell)} + \varepsilon \bar{b}_{11}^{(\ell)} \mu^{d-2\ell} \right) \mu^\ell = \delta_1, \tag{3.16}$$

$$\sum_{\ell=0}^{(d-1)/2} \rho_\ell \left(b_{21}^{(\ell)} + \varepsilon \bar{b}_{12}^{(\ell)} \mu^{d-2\ell} \right) \mu^\ell = \delta_2. \tag{3.17}$$

We observe that:

1. $b_{22}^{(\ell)}$ for all ℓ do not show up. Thus must $b_{22}^{(\ell)} = 0$ for all ℓ by (2.8).
2. Equations (3.13) and (3.14) are decoupled. Thus they can be solved separately for optimal solutions that minimize

$$\sum_{\ell=0}^{d/2} |b_{11}^{(\ell)}|^2, \quad 2|b_{12}^{(d/2)}|^2 + \sum_{\ell=0}^{d/2-1} \left(|b_{12}^{(\ell)}|^2 + |b_{21}^{(\ell)}|^2 \right), \tag{3.18}$$

respectively, to arrive at optimal solutions to (3.12) in the sense of (2.8). Finally Δ_F^2 is the sum of the optimized values of the two expressions in (3.18). This is for even d.

3. Equations (3.16) and (3.17) are decoupled, too. Thus they can also be solved separately for optimal solutions that minimize

$$\sum_{\ell=0}^{(d-1)/2} |b_{11}^{(\ell)}|^2, \quad \sum_{\ell=0}^{(d-1)/2} \left(|b_{12}^{(\ell)}|^2 + |b_{21}^{(\ell)}|^2\right), \tag{3.19}$$

respectively, to arrive at optimal solutions to (3.15) in the sense of (2.8). Finally Δ_F^2 is the sum of the optimized values of the two expressions in (3.19). This is for odd d.

We shall now solve each of (3.13), (3.14), (3.16), and (3.17) in terms of four lemmas. For this purpose, write

$$\mu = |\mu| e^{\iota\theta}, \quad c_\ell = \cos(\ell\theta), \quad s_\ell = \sin(\ell\theta). \tag{3.20}$$

Then by Lemma 2.1 (i), we have

$$\begin{aligned}
\begin{pmatrix} \Re\left(\left[b_{11}^{(\ell)} + \varepsilon \bar{b}_{11}^{(\ell)} \mu^{d-2\ell}\right] \mu^\ell\right) \\ \Im\left(\left[b_{11}^{(\ell)} + \varepsilon \bar{b}_{11}^{(\ell)} \mu^{d-2\ell}\right] \mu^\ell\right) \end{pmatrix} &= |\mu|^\ell \begin{pmatrix} c_\ell & -s_\ell \\ s_\ell & c_\ell \end{pmatrix} \begin{pmatrix} \Re(b_{11}^{(\ell)} + \varepsilon \bar{b}_{11}^{(\ell)} \mu^{d-2\ell}) \\ \Im(b_{11}^{(\ell)} + \varepsilon \bar{b}_{11}^{(\ell)} \mu^{d-2\ell}) \end{pmatrix} \\
&= |\mu|^\ell \begin{pmatrix} c_\ell & -s_\ell \\ s_\ell & c_\ell \end{pmatrix} \left[I_2 + \varepsilon |\mu|^{d-2\ell} \begin{pmatrix} c_{d-2\ell} & s_{d-2\ell} \\ s_{d-2\ell} & -c_{d-2\ell} \end{pmatrix} \right] \begin{pmatrix} \Re(b_{11}^{(\ell)}) \\ \Im(b_{11}^{(\ell)}) \end{pmatrix}.
\end{aligned} \tag{3.21}$$

Lemma 3.1 *Equation* (3.13) *which is for even d has a solution if and only if either $|\mu| \neq 1$ or $|\mu| = 1$ and $z^{\mathrm{H}} r / (\sqrt{\varepsilon} \mu^{d/2}) \in \mathbb{R}$. Otherwise it is inconsistent. When* (3.13) *has a solution, its optimal solution that minimizes $\sum_{\ell=0}^{d/2} |b_{11}^{(\ell)}|^2$ is given by, for $0 \le \ell \le d/2 - 1$,*

$$\begin{aligned}
\begin{pmatrix} \Re(b_{11}^{(\ell)}) \\ \Im(b_{11}^{(\ell)}) \end{pmatrix} &= \rho_\ell \begin{pmatrix} c_{d/2-\ell} & s_{d/2-\ell} \\ s_{d/2-\ell} & -c_{d/2-\ell} \end{pmatrix} \begin{pmatrix} \dfrac{|\mu|^\ell + \varepsilon |\mu|^{d-\ell}}{\xi_{\text{even}} + \varepsilon \zeta_{\text{even}}} & \\ & \dfrac{|\mu|^\ell - \varepsilon |\mu|^{d-\ell}}{\xi_{\text{even}} - \varepsilon \zeta_{\text{even}}} \end{pmatrix} \\
&\quad \times \begin{pmatrix} c_{d/2} & s_{d/2} \\ s_{d/2} & -c_{d/2} \end{pmatrix} \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix},
\end{aligned} \tag{3.22}$$

and[2]

$$\varepsilon\sqrt{\varepsilon}b_{11}^{(d/2)} = \frac{\rho_{d/2}|\mu|^{d/2}}{\xi_{\text{even}}+\zeta_{\text{even}}}(c_{d/2}\quad s_{d/2})J_\varepsilon^\top\begin{pmatrix}\Re(\delta_1)\\ \Im(\delta_1)\end{pmatrix},\quad J_\varepsilon = \frac{1}{2}\begin{pmatrix}1+\varepsilon & \varepsilon-1\\ 1-\varepsilon & 1+\varepsilon\end{pmatrix}, \tag{3.23}$$

and the optimal value[3]

$$\sum_{\ell=0}^{d/2}|b_{11}^{(\ell)}|^2 = \begin{cases}\dfrac{|\Re(e^{-\iota\theta d/2}\,\delta_1)|^2}{\Phi_{\text{even}}}+\dfrac{|\Im(e^{-\iota\theta d/2}\,\delta_1)|^2}{\phi_{\text{even}}} & \text{for } \varepsilon=1,\\ \dfrac{|\Re(e^{-\iota\theta d/2}\,\delta_1)|^2}{\phi_{\text{even}}}+\dfrac{|\Im(e^{-\iota\theta d/2}\,\delta_1)|^2}{\Phi_{\text{even}}} & \text{for } \varepsilon=-1\end{cases} \tag{3.24}$$

$$\leq \frac{|\delta_1|^2}{\phi_{\text{even}}}. \tag{3.25}$$

When $|\mu| = 1$ *and* $z^{\mathrm{H}}r/(\sqrt{\varepsilon}\mu^{d/2}) \in \mathbb{R}$*,* (3.22)*,* (3.23)*, and* (3.24) *to, for* $0 \leq \ell \leq d/2-1$*,*

$$\begin{pmatrix}\Re(b_{11}^{(\ell)})\\ \Im(b_{11}^{(\ell)})\end{pmatrix} = \frac{2\rho_\ell}{\Phi_{\text{even}}}J_\varepsilon\begin{pmatrix}c_{d/2-\ell}\\ s_{d/2-\ell}\end{pmatrix}\frac{\delta_1}{\sqrt{\varepsilon}\mu^{d/2}},\quad \varepsilon\sqrt{\varepsilon}b_{11}^{(d/2)} = \frac{\rho_{d/2}}{\Phi_{\text{even}}}\frac{\delta_1}{\sqrt{\varepsilon}\mu^{d/2}}, \tag{3.26}$$

and the optimal value

$$\sum_{\ell=0}^{d/2}|b_{11}^{(\ell)}|^2 = \frac{|\delta_1|^2}{\Phi_{\text{even}}}. \tag{3.27}$$

Proof Equation (3.13) needs to be split into two equations according to the real and imaginary parts because $\sqrt{\varepsilon}b_{11}^{(d/2)}$ is real and both $b_{11}^{(\ell)}$ and $\bar{b}_{11}^{(\ell)}$ for $\ell = 0, 1, \ldots, d/2-1$ appear. By (3.21), we see that (3.13) is equivalently to

$$(Z_0\ \ldots\ Z_{d/2-1}\ \ Z_{d/2})\begin{pmatrix}\Re(b_{11}^{(0)})\\ \Im(b_{11}^{(0)})\\ \vdots\\ \Re(b_{11}^{(d/2-1)})\\ \Im(b_{11}^{(d/2-1)})\\ \varepsilon\sqrt{\varepsilon}b_{11}^{(d/2)}\end{pmatrix} = \begin{pmatrix}\Re(\delta_1)\\ \Im(\delta_1)\end{pmatrix}, \tag{3.28}$$

[2] These strange looking formulas are the artifacts of putting both cases $\varepsilon = \pm 1$ together. In fact, $J_1 = I_2$ and $J_{-1} = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$, and $\varepsilon\sqrt{\varepsilon}b_{11}^{(d/2)}$ is simply $b_{11}^{(d/2)}$ when $\varepsilon = 1$ (for which $b_{11}^{(d/2)}$ is real), and $\Im(b_{11}^{(d/2)})$ when $\varepsilon = -1$ (for which $b_{11}^{(d/2)}$ is pure imaginary).

[3] $e^{-\iota\theta d/2}\,\delta_1$ can also be expressed as $\left(\dfrac{|\mu|}{\mu}\right)^{d/2}\dfrac{z^{\mathrm{H}}r}{\|z\|_2^2}$.

where for $\ell = 0, 1, \ldots, d/2 - 1$

$$Z_\ell = \rho_\ell |\mu|^\ell \begin{pmatrix} c_\ell & -s_\ell \\ s_\ell & c_\ell \end{pmatrix} \left[I_2 + \varepsilon |\mu|^{d-2\ell} \begin{pmatrix} c_{d-2\ell} & s_{d-2\ell} \\ s_{d-2\ell} & -c_{d-2\ell} \end{pmatrix} \right], \tag{3.29}$$

and

$$Z_{d/2} = \rho_{d/2} |\mu|^{d/2} J_\varepsilon \begin{pmatrix} c_{d/2} \\ s_{d/2} \end{pmatrix}. \tag{3.30}$$

Set $Z = (Z_0, \ldots, Z_{d/2-1} \;\; Z_{d/2}) \in \mathbb{R}^{2\times(d+1)}$. It can be computed, by Lemma 2.1 (ii), that

$$\begin{aligned} Z_\ell Z_\ell^\top &= \rho_\ell^2 |\mu|^{2\ell} \begin{pmatrix} c_\ell & -s_\ell \\ s_\ell & c_\ell \end{pmatrix} \left[I_2 + \varepsilon |\mu|^{d-2\ell} \begin{pmatrix} c_{d-2\ell} & s_{d-2\ell} \\ s_{d-2\ell} & -c_{d-2\ell} \end{pmatrix} \right]^2 \begin{pmatrix} c_\ell & -s_\ell \\ s_\ell & c_\ell \end{pmatrix}^\top \\ &= \rho_\ell^2 \left[\left(|\mu|^{2\ell} + |\mu|^{2(d-\ell)} \right) I_2 + 2\varepsilon |\mu|^d \begin{pmatrix} c_d & s_d \\ s_d & -c_d \end{pmatrix} \right] \quad \text{for } 0 \le \ell \le d/2 - 1, \\ Z_{d/2} Z_{d/2}^\top &= \rho_{d/2}^2 |\mu|^d J_\varepsilon \begin{pmatrix} c_{d/2} \\ s_{d/2} \end{pmatrix} \begin{pmatrix} c_{d/2} \\ s_{d/2} \end{pmatrix}^\top J_\varepsilon^\top \\ &= \frac{\rho_{d/2}^2 |\mu|^d}{2} \left[I_2 + \varepsilon \begin{pmatrix} c_d & s_d \\ s_d & -c_d \end{pmatrix} \right]. \end{aligned}$$

Therefore

$$ZZ^\top = \sum_{\ell=0}^{d/2} Z_\ell Z_\ell^\top = \xi_{\text{even}} I_2 + \varepsilon \zeta_{\text{even}} \begin{pmatrix} c_d & s_d \\ s_d & -c_d \end{pmatrix}. \tag{3.31}$$

The eigenvalues of $\begin{pmatrix} c_d & s_d \\ s_d & -c_d \end{pmatrix}$ are ± 1. In fact, its eigendecomposition can be explicitly written out as

$$\begin{pmatrix} c_d & s_d \\ s_d & -c_d \end{pmatrix} = \begin{pmatrix} c_{d/2} & s_{d/2} \\ s_{d/2} & -c_{d/2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} c_{d/2} & s_{d/2} \\ s_{d/2} & -c_{d/2} \end{pmatrix}. \tag{3.32}$$

So the eigenvalues of $ZZ^\top$ are Φ_{even} and ϕ_{even}, and moreover, we have an explicit eigendecomposition for $ZZ^\top$:

$$ZZ^\top = \begin{pmatrix} c_{d/2} & s_{d/2} \\ s_{d/2} & -c_{d/2} \end{pmatrix} \begin{pmatrix} \xi_{\text{even}} + \varepsilon \zeta_{\text{even}} & \\ & \xi_{\text{even}} - \varepsilon \zeta_{\text{even}} \end{pmatrix} \begin{pmatrix} c_{d/2} & s_{d/2} \\ s_{d/2} & -c_{d/2} \end{pmatrix}. \tag{3.33}$$

If $|\mu| \neq 1$, then the smaller eigenvalue $\phi_{\text{even}} > 0$ and thus $ZZ^\top$ is invertible. The optimal solution that minimizes of $\sum_{\ell=0}^{d/2} |b_{11}^{(\ell)}|^2$ is

$$\begin{pmatrix} \Re(b_{11}^{(0)}) \\ \Im(b_{11}^{(0)}) \\ \vdots \\ \Re(b_{11}^{(d/2-1)}) \\ \Im(b_{11}^{(d/2-1)}) \\ \varepsilon\sqrt{\varepsilon} b_{11}^{(d/2)} \end{pmatrix} = Z^\dagger \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix} = Z^\top (ZZ^\top)^{-1} \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix}. \tag{3.34}$$

To arrive at (3.22), (3.23), and (3.24), we notice from (3.34) that

$$\begin{pmatrix} \Re(b_{11}^{(\ell)}) \\ \Im(b_{11}^{(\ell)}) \end{pmatrix} = Z_\ell^\top (ZZ^\top)^{-1} \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix}, \quad \varepsilon\sqrt{\varepsilon} b_{11}^{(d/2)} = Z_{d/2}^\top (ZZ^\top)^{-1} \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix}. \tag{3.35}$$

It can be verified that for $0 \leq \ell \leq d/2 - 1$

$$Z_\ell = \rho_\ell |\mu|^\ell \begin{pmatrix} c_{d/2} & s_{d/2} \\ s_{d/2} & -c_{d/2} \end{pmatrix} \begin{pmatrix} 1 + \varepsilon|\mu|^{d-2\ell} & \\ & 1 - \varepsilon|\mu|^{d-2\ell} \end{pmatrix} \begin{pmatrix} c_{d/2-\ell} & s_{d/2-\ell} \\ s_{d/2-\ell} & -c_{d/2-\ell} \end{pmatrix}. \tag{3.36}$$

Equations (3.22) and (3.23) are the consequences of (3.30), (3.33), (3.35), and (3.36), upon noticing

$$\begin{pmatrix} c_{d/2} & s_{d/2} \\ s_{d/2} & -c_{d/2} \end{pmatrix} \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix} = \begin{pmatrix} \Re(e^{-\iota\theta d/2}\,\delta_1) \\ -\Im(e^{-\iota\theta d/2}\,\delta_1) \end{pmatrix},$$
$$(c_{d/2} \quad s_{d/2}) J_\varepsilon^\top \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix} = \begin{cases} \Re(e^{-\iota\theta d/2}\,\delta_1) & \text{for } \varepsilon = 1, \\ -\Im(e^{-\iota\theta d/2}\,\delta_1) & \text{for } \varepsilon = -1. \end{cases}$$

Finally we have, by (3.22) and (3.23), the optimal value

$$\sum_{\ell=0}^{d/2} |b_{11}^{(\ell)}|^2 = \sum_{\ell=0}^{d/2-1} \rho_\ell^2 \left[\left(\frac{|\mu|^\ell + \varepsilon|\mu|^{d-\ell}}{\xi_{\text{even}} + \varepsilon\zeta_{\text{even}}} \right)^2 |\Re(e^{-\iota\theta d/2}\,\delta_1)|^2 + \left(\frac{|\mu|^\ell - \varepsilon|\mu|^{d-\ell}}{\xi_{\text{even}} - \varepsilon\zeta_{\text{even}}} \right)^2 |\Im(e^{-\iota\theta d/2}\,\delta_1)|^2 \right] + \left(\frac{\rho_{d/2}|\mu|^{d/2}}{\xi_{\text{even}} + \zeta_{\text{even}}} \right)^2 \left| (c_{d/2} \quad s_{d/2}) J_\varepsilon^\top \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix} \right|^2$$

which leads to (3.24).

If, however, $|\mu| = 1$, then $\operatorname{rank}(Z) = 1$ and (3.28) is not solvable unless $(\Re(\delta_1)\ \ \Im(\delta_1))^\top$ is parallel to $J_\varepsilon^\top (c_{d/2}\ \ s_{d/2})^\top$, or equivalently

$$\frac{\delta_1}{\sqrt{\varepsilon}\mu^{d/2}} = \frac{z^{\mathrm{H}} r}{\sqrt{\varepsilon}\mu^{d/2}\|z\|_2^2} \in \mathbb{R}. \tag{3.37}$$

When it does, the optimal solution can be gotten from either equation from the first or the second component in (3.28). We shall do so now. When $|\mu| = 1$, we have for $\ell = 0, 1, \ldots, d/2 - 1$

$$\begin{aligned} Z_\ell &= \rho_\ell \begin{pmatrix} c_\ell & -s_\ell \\ s_\ell & c_\ell \end{pmatrix} \left[I_2 + \varepsilon \begin{pmatrix} c_{d-2\ell} & s_{d-2\ell} \\ s_{d-2\ell} & -c_{d-2\ell} \end{pmatrix} \right] \\ &= \rho_\ell \left[\begin{pmatrix} c_\ell & -s_\ell \\ s_\ell & c_\ell \end{pmatrix} + \varepsilon \begin{pmatrix} c_{d-\ell} & s_{d-\ell} \\ s_{d-\ell} & -c_{d-\ell} \end{pmatrix} \right] \\ &= 2\rho_\ell J_\varepsilon \begin{pmatrix} c_{d/2} \\ s_{d/2} \end{pmatrix} \begin{pmatrix} c_{d/2-\ell} \\ s_{d/2-\ell} \end{pmatrix}^\top J_\varepsilon^\top, \end{aligned} \tag{3.38}$$

and

$$Z_{d/2} = \rho_{d/2} J_\varepsilon \begin{pmatrix} c_{d/2} \\ s_{d/2} \end{pmatrix}.$$

To see (3.26) and (3.27), we take the first equation in (3.28) for example. With $|\mu| = 1$, the first equation is

$$\begin{aligned} &\sum_{\ell=0}^{d/2-1} \rho_\ell \left[(c_\ell + \varepsilon c_{d-\ell}) \Re(b_{11}^{(\ell)}) + (-s_\ell + \varepsilon s_{d-\ell}) \Im(b_{11}^{(\ell)}) \right] \\ &\quad + \frac{1}{2} \rho_{d/2} \left[(1+\varepsilon) c_{d/2} - (1-\varepsilon) s_{d/2} \right] \varepsilon \sqrt{\varepsilon} b_{11}^{d/2} = \Re(\delta_1). \end{aligned} \tag{3.39}$$

For $\varepsilon = 1$, it is

$$2c_{d/2} \sum_{\ell=0}^{d/2-1} \rho_\ell \left[c_{d/2-\ell} \Re\left(b_{11}^{(\ell)}\right) + s_{d-\ell/2} \Im\left(b_{11}^{(\ell)}\right) \right] + \rho_{d/2} c_{d/2}\, \varepsilon \sqrt{\varepsilon} b_{11}^{d/2} = \Re(\delta_1)$$

which, under the condition that $c_{d/2} \neq 0$, leads to (3.26) and (3.27) upon noticing that $\Re(\delta_1)/c_{d/2} = \delta_1/\mu^{d/2}$ because of (3.37). If $c_{d/2} = 0$, the second equation in (3.28) will lead to the same conclusion for the case. The case for $\varepsilon = -1$ is similar. □

Remark 3.2 Inequality (3.25) becomes an equality when $(\Re(\delta_1)\ \ \Im(\delta_1))^\top$ is parallel to the eigenvector of $\begin{pmatrix} c_d & s_d \\ s_d & -c_d \end{pmatrix}$ associated with its eigenvalue $-\varepsilon$. As a consequence, (3.11) can become an equality, too, for even d.

Remark 3.3 Equation (3.24) reveals what contributes the most to the structured backward error, while (3.25) fails to do at the price of being simple. That is that not all of $|\delta_1|$ contributes equally in proportional to $1/\sqrt{\phi_{\rm even}}$ to the structured backward error. In fact only either $|\Re(e^{-\iota\theta d/2}\,\delta_1)|$ or $|\Im(e^{-\iota\theta d/2}\,\delta_1)|$, depending on what ε is, does. Conceivably, there are circumstances where the extreme tininess of either $|\Re(e^{-\iota\theta d/2}\,\delta_1)|$ or $|\Im(e^{-\iota\theta d/2}\,\delta_1)|$ may offset the tininess of $\sqrt{\phi_{\rm even}}$ to render reasonably tiny structured backward error, much smaller than (3.25) would indicate. Similar comment applies to part of the development in Lemma 3.3, too.

Lemma 3.2 *Equation* (3.14) *which is for even d always has a solution. Its optimal solution that minimizes* $2|b_{12}^{(d/2)}|^2+\sum_{\ell=0}^{d/2-1}\left(|b_{12}^{(\ell)}|^2+|b_{21}^{(\ell)}|^2\right)$ *is given by*

$$\sqrt{2}\,b_{12}^{(d/2)}=\varepsilon\frac{\rho_{d/2}\,\mu^{d/2}\bar{\delta}_2/\sqrt{2}}{\Psi_{\rm even}},\tag{3.40}$$

$$b_{12}^{(\ell)}=\varepsilon\frac{\rho_\ell\,\mu^{d-\ell}\,\bar{\delta}_2}{\Psi_{\rm even}},\quad b_{21}^{(\ell)}=\frac{\rho_\ell\,\bar{\mu}^{\ell}\,\delta_2}{\Psi_{\rm even}}\quad \textit{for } \ell=0,1,\ldots,d/2-1\tag{3.41}$$

satisfying

$$2\left|b_{12}^{(d/2)}\right|^2+\sum_{\ell=0}^{d/2-1}\left(\left|b_{12}^{(\ell)}\right|^2+\left|b_{21}^{(\ell)}\right|^2\right)=\frac{|\delta_2|^2}{\Psi_{\rm even}}.\tag{3.42}$$

Proof $\Psi_{\rm even}>0$ always since by assumption all $\rho_\ell>0$. It can be verified that $b_{ij}^{(\ell)}$'s given by (3.40) and (3.41) satisfy (3.14). On the other hand, (3.14) implies, by the Cauchy-Schwarz inequality,

$$|\delta_2|^2\le\Psi_{\rm even}\left[2\left|b_{12}^{(d/2)}\right|^2+\sum_{\ell=0}^{d/2-1}\left(\left|b_{12}^{(\ell)}\right|^2+\left|b_{21}^{(\ell)}\right|^2\right)\right]$$

for any solution $\{b_{ij}^{(\ell)}\}$ to (3.14). □

Lemma 3.3 *Equation* (3.16) *which is for odd d has a solution if and only if either $|\mu|\neq 1$ or $|\mu|=1$ and $z^{\rm H}r/(\sqrt{\varepsilon}\mu^{d/2})\in\mathbb{R}$. Otherwise it is inconsistent. When* (3.16) *has a solution, its optimal solution that minimizes* $\sum_{\ell=0}^{(d-1)/2}|b_{11}^{(\ell)}|^2$ *is given by, for* $0\le\ell\le(d-1)/2-1$,

$$\begin{pmatrix}\Re(b_{11}^{(\ell)})\\ \Im(b_{11}^{(\ell)})\end{pmatrix}=\rho_\ell\begin{pmatrix}c_{d/2-\ell} & s_{d/2-\ell}\\ s_{d/2-\ell} & -c_{d/2-\ell}\end{pmatrix}\begin{pmatrix}\dfrac{|\mu|^\ell+\varepsilon|\mu|^{d-\ell}}{\xi_{\rm odd}+\varepsilon\zeta_{\rm odd}} & \\ & \dfrac{|\mu|^\ell-\varepsilon|\mu|^{d-\ell}}{\xi_{\rm odd}-\varepsilon\zeta_{\rm odd}}\end{pmatrix}$$
$$\times\begin{pmatrix}c_{d/2} & s_{d/2}\\ s_{d/2} & -c_{d/2}\end{pmatrix}\begin{pmatrix}\Re(\delta_1)\\ \Im(\delta_1)\end{pmatrix},\tag{3.43}$$

and the optimal value

$$\sum_{\ell=0}^{(d-1)/2} \left|b_{11}^{(\ell)}\right|^2 = \begin{cases} \dfrac{|\Re(e^{-\iota\theta d/2}\,\delta_1)|^2}{\Phi_{\text{odd}}} + \dfrac{|\Im(e^{-\iota\theta d/2}\,\delta_1)|^2}{\phi_{\text{odd}}} & \textit{for } \varepsilon = 1, \\ \dfrac{|\Re(e^{-\iota\theta d/2}\,\delta_1)|^2}{\phi_{\text{odd}}} + \dfrac{|\Im(e^{-\iota\theta d/2}\,\delta_1)|^2}{\Phi_{\text{odd}}} & \textit{for } \varepsilon = -1 \end{cases} \tag{3.44}$$

$$\leq \frac{|\delta_1|^2}{\phi_{\text{odd}}}. \tag{3.45}$$

When $|\mu| = 1$ *and* $z^{\mathrm{H}} r/(\sqrt{\varepsilon}\mu^{d/2}) \in \mathbb{R}$, (3.43) *and* (3.44) *can be significantly simplified to, for* $0 \leq \ell \leq (d-1)/2 - 1$,

$$\begin{pmatrix} \Re\left(b_{11}^{(\ell)}\right) \\ \Im\left(b_{11}^{(\ell)}\right) \end{pmatrix} = \frac{2\rho_\ell}{\Phi_{\text{odd}}} J_\varepsilon \begin{pmatrix} c_{d/2-\ell} \\ s_{d/2-\ell} \end{pmatrix} \frac{\delta_1}{\sqrt{\varepsilon}\mu^{d/2}}, \tag{3.46}$$

and the optimal value

$$\sum_{\ell=0}^{(d-1)/2} \left|b_{11}^{(\ell)}\right|^2 = \frac{|\delta_1|^2}{\Phi_{\text{odd}}}. \tag{3.47}$$

Proof Similarly to the proof of Lemma 3.1 above, (3.16) is equivalently to

$$(Z_0, \ldots, Z_{(d-1)/2}) \begin{pmatrix} \Re\left(b_{11}^{(0)}\right) \\ \Im\left(b_{11}^{(0)}\right) \\ \vdots \\ \Re\left(b_{11}^{((d-1)/2)}\right) \\ \Im\left(b_{11}^{((d-1)/2)}\right) \end{pmatrix} = \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix}, \tag{3.48}$$

where Z_ℓ are as in (3.29). Let $Z = (Z_0, \ldots, Z_{(d-1)/2})$. Then, as in the proof of Lemma 3.1,

$$\begin{aligned} ZZ^\top &= \xi_{\text{odd}} I_2 + \varepsilon\zeta_{\text{odd}} \begin{pmatrix} c_d & s_d \\ s_d & -c_d \end{pmatrix} \\ &= \begin{pmatrix} c_{d/2} & s_{d/2} \\ s_{d/2} & -c_{d/2} \end{pmatrix} \begin{pmatrix} \xi_{\text{odd}} + \varepsilon\zeta_{\text{odd}} & \\ & \xi_{\text{odd}} - \varepsilon\zeta_{\text{odd}} \end{pmatrix} \begin{pmatrix} c_{d/2} & s_{d/2} \\ s_{d/2} & -c_{d/2} \end{pmatrix}. \end{aligned}$$

The eigenvalues of $ZZ^\top$ are Φ_{odd} and ϕ_{odd}. If $|\mu| \neq 1$, then the smaller eigenvalue $\phi_{\text{odd}} > 0$ and thus $ZZ^\top$ is invertible. The optimal solution that minimizes

$\sum_{\ell=0}^{(d-1)/2} |b_{11}^{(\ell)}|^2$ is, for $\ell = 0, 1, \ldots, (d-1)/2 - 1$,

$$\begin{pmatrix} \Re\left(b_{11}^{(\ell)}\right) \\ \Im\left(b_{11}^{(\ell)}\right) \end{pmatrix} = Z_\ell^\top (ZZ^\top)^{-1} \begin{pmatrix} \Re(\delta_1) \\ \Im(\delta_1) \end{pmatrix} \tag{3.49}$$

which lead to (3.43) and (3.44).

If, however, $|\mu| = 1$, then $\text{rank}(Z) = 1$ and (3.48) is not solvable unless $(\Re(\delta_1)\ \ \Im(\delta_1))^\top$ is parallel to $J_\varepsilon^\top (c_{d/2}\ \ s_{d/2})^\top$, or equivalently (3.37) holds. When it does, the optimal solution can be gotten similarly as in the proof of Lemma 3.1. □

Remark 3.4 Inequality (3.45) becomes an equality when $(\Re(\delta_1)\ \ \Im(\delta_1))^\top$ is parallel to the eigenvector of $\begin{pmatrix} c_d & s_d \\ s_d & -c_d \end{pmatrix}$ associated with its eigenvalue $-\varepsilon$. As a consequence, (3.11) can become an equality, too, for odd d.

Lemma 3.4 *Equation* (3.17) *which is for odd d always has a solution. Its optimal solution that minimizes* $\sum_{\ell=0}^{(d-1)/2} \left(|b_{12}^{(\ell)}|^2 + |b_{21}^{(\ell)}|^2\right)$ *is given by*

$$b_{12}^{(\ell)} = \varepsilon \frac{\rho_\ell\, \mu^{d-\ell} \bar{\delta}_2}{\Psi_{\text{odd}}}, \quad b_{21}^{(\ell)} = \frac{\rho_\ell\, \bar{\mu}^\ell\, \delta_2}{\Psi_{\text{odd}}} \quad \textit{for } \ell = 0, 1, \ldots, (d-1)/2 \tag{3.50}$$

satisfying

$$\sum_{\ell=0}^{(d-1)/2} \left(\left|b_{12}^{(\ell)}\right|^2 + \left|b_{21}^{(\ell)}\right|^2\right) = \frac{|\delta_2|^2}{\Psi_{\text{odd}}}. \tag{3.51}$$

Proof $\Psi_{\text{odd}} > 0$ always since by assumption all $\rho_\ell > 0$. It can be verified that $b_{ij}^{(\ell)}$'s given by (3.50) satisfy (3.17). On the other hand, (3.17) implies, by the Cauchy–Schwarz inequality,

$$|\delta_2|^2 \leq \Psi_{\text{odd}} \left[\sum_{\ell=0}^{(d-1)/2-1} \left(\left|b_{12}^{(\ell)}\right|^2 + \left|b_{21}^{(\ell)}\right|^2\right)\right]$$

for any solution $\{b_{ij}^{(\ell)}\}$ to (3.17). □

Remark 3.5 Our solutions to (3.13) and (3.16) by Lemmas 3.1 and 3.3 are all that are needed for dealing with the scalar polynomial case, i.e., when $n = 1$.

4 Structured backward error for $\star = \top$

Theorem 4.1 is the main result of this section. We first define a few parameters in term of a given approximate eigenpair $\{\mu, z\}$ of PPEP (1.4): for even d,

$$\Phi_{\rm even} \stackrel{\rm def}{=} \sum_{\ell=0}^{d/2-1} \rho_\ell^2 \left|\mu^\ell + \varepsilon\mu^{d-\ell}\right|^2 + \sigma_\varepsilon \rho_{d/2}^2 |\mu|^d, \tag{4.1}$$

$$\Psi_{\rm even} \stackrel{\rm def}{=} \sum_{\ell=0}^{d/2-1} \rho_\ell^2 \left(|\mu|^{2\ell} + |\mu|^{2(d-\ell)}\right) + \rho_{d/2}^2 |\mu|^d/2, \tag{4.2}$$

and for odd d,

$$\Phi_{\rm odd} \stackrel{\rm def}{=} \sum_{\ell=0}^{(d-1)/2} \rho_\ell^2 \left|\mu^\ell + \varepsilon\mu^{d-\ell}\right|^2, \tag{4.3}$$

$$\Psi_{\rm odd} \stackrel{\rm def}{=} \sum_{\ell=0}^{(d-1)/2} \rho_\ell^2 \left(|\mu|^{2\ell} + |\mu|^{2(d-\ell)}\right). \tag{4.4}$$

Throughout this section, these assignments (4.1)–(4.4) are assumed.

Theorem 4.1 *Let $\{\mu, z\}$ be a given approximate eigenpair of PPEP* (1.4)*. Suppose $\star = \top$ and $\varepsilon = \pm 1$ in* (1.5)*, and δ_1 and δ_2 are as in* (2.11) *with $\widehat{r} = \bar{r}$ which is defined in* (2.1)*. Let*

$$\Phi = \begin{cases} \Phi_{\rm even} & \text{for even } d, \\ \Phi_{\rm odd}, & \text{for odd } d, \end{cases} \qquad \Psi = \begin{cases} \Psi_{\rm even} & \text{for even } d, \\ \Psi_{\rm odd}, & \text{for odd } d. \end{cases}$$

For the structured backward error $\Delta_{\rm F}$, we have

1. *If d is odd, $\mu = -\varepsilon$, or if d is even, $\mu = \pm 1$, and $\varepsilon = -1$, then*

$$\Delta_{\rm F} = \begin{cases} +\infty, & \text{if } z^{\rm H}\bar{r} \neq 0, \\ |\delta_2|/\sqrt{\Psi}, & \text{if } z^{\rm H}\bar{r} = 0. \end{cases}$$

2. *If d is odd, $\mu \neq -\varepsilon$, or if d is even, $\mu \neq \pm 1$, or if d is even, $\varepsilon = 1$, then*

$$\Delta_{\rm F} = \sqrt{\frac{|\delta_1|^2}{\Phi} + \frac{|\delta_2|^2}{\Psi}}. \tag{4.5}$$

The proof of this theorem also rests on solving (2.7). It is done by Lemmas 4.1–4.4 below. Recall (2.15), (2.16), (2.17), and Theorem 2.1, and keep in mind that $\star = \top$, $\varepsilon = \pm 1$, and $\|\cdot\| = \|\cdot\|_{\rm F}$ in (2.8).

For even d, we need to solve

$$\left[\sum_{\ell=0}^{d/2-1} \rho_\ell \left(\begin{pmatrix} b_{11}^{(\ell)} & b_{12}^{(\ell)} \\ b_{21}^{(\ell)} & b_{22}^{(\ell)} \end{pmatrix} + \varepsilon \begin{pmatrix} b_{11}^{(\ell)} & b_{21}^{(\ell)} \\ b_{12}^{(\ell)} & b_{22}^{(\ell)} \end{pmatrix} \mu^{d-2\ell} \right) \mu^\ell \right.$$
$$\left. + \rho_{d/2} \begin{pmatrix} \sigma_\varepsilon b_{11}^{(d/2)} & b_{12}^{(d/2)} \\ \varepsilon b_{12}^{(d/2)} & \sigma_\varepsilon b_{22}^{(d/2)} \end{pmatrix} \mu^{d/2} \right] \begin{pmatrix} \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} \bar{\gamma} \\ \bar{\beta} \end{pmatrix}, \tag{4.6}$$

where $\sigma_\varepsilon = \frac{1+\varepsilon}{2}$, i.e., σ_ε equals to 1 for $\varepsilon = 1$ and -1 for $\varepsilon = -1$. Componentwise, it gives

$$\sum_{\ell=0}^{d/2-1} \rho_\ell \left(b_{11}^{(\ell)} + \varepsilon b_{11}^{(\ell)} \mu^{d-2\ell} \right) \mu^\ell + \rho_{d/2} \sigma_\varepsilon b_{11}^{(d/2)} \mu^{d/2} = \delta_1, \tag{4.7}$$

$$\sum_{\ell=0}^{d/2-1} \rho_\ell \left(b_{21}^{(\ell)} + \varepsilon b_{12}^{(\ell)} \mu^{d-2\ell} \right) \mu^\ell + \varepsilon \rho_{d/2} b_{12}^{(d/2)} \mu^{d/2} = \delta_2. \tag{4.8}$$

For odd d, we need to solve

$$\left[\sum_{\ell=0}^{(d-1)/2} \rho_\ell \left(\begin{pmatrix} b_{11}^{(\ell)} & b_{12}^{(\ell)} \\ b_{21}^{(\ell)} & b_{22}^{(\ell)} \end{pmatrix} + \varepsilon \begin{pmatrix} b_{11}^{(\ell)} & b_{21}^{(\ell)} \\ b_{12}^{(\ell)} & b_{22}^{(\ell)} \end{pmatrix} \mu^{d-2\ell} \right) \mu^\ell \right] \begin{pmatrix} \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} \bar{\gamma} \\ \bar{\beta} \end{pmatrix}. \tag{4.9}$$

Componentwise, it gives

$$\sum_{\ell=0}^{(d-1)/2} \rho_\ell \left(b_{11}^{(\ell)} + \varepsilon b_{11}^{(\ell)} \mu^{d-2\ell} \right) \mu^\ell = \delta_1, \tag{4.10}$$

$$\sum_{\ell=0}^{(d-1)/2} \rho_\ell \left(b_{21}^{(\ell)} + \varepsilon b_{12}^{(\ell)} \mu^{d-2\ell} \right) \mu^\ell = \delta_2. \tag{4.11}$$

We observe that:

1. $b_{22}^{(\ell)}$ for all ℓ do not show up. Thus must $b_{22}^{(\ell)} = 0$ for all ℓ by (2.8).
2. Equations (4.7) and (4.8) are decoupled. Thus they can be solved separately for optimal solutions that minimize

$$\sum_{\ell=0}^{d/2} |b_{11}^{(\ell)}|^2, \quad 2|b_{12}^{(d/2)}|^2 + \sum_{\ell=0}^{d/2-1} \left(|b_{12}^{(\ell)}|^2 + |b_{21}^{(\ell)}|^2 \right), \tag{4.12}$$

   respectively, to arrive at optimal solutions to (4.6) in the sense of (2.8). Finally Δ_{F}^2 is the sum of the optimized values of the two expressions in (4.12). This is for even d.

3. Equations (4.10) and (4.11) are decoupled, too. Thus they can also be solved separately for optimal solutions that minimize

$$\sum_{\ell=0}^{(d-1)/2} \left|b_{11}^{(\ell)}\right|^2, \quad \sum_{\ell=0}^{(d-1)/2} \left(\left|b_{12}^{(\ell)}\right|^2 + \left|b_{21}^{(\ell)}\right|^2\right), \tag{4.13}$$

respectively, to arrive at optimal solutions to (4.9) in the sense of (2.8). Finally Δ_F^2 is the sum of the optimized values of the two expressions in (4.13). This is for odd d.

We shall now solve each of (4.7), (4.8), (4.10), and (4.11) in terms of four lemmas.

Lemma 4.1 *Equation* (4.7) *which is for even d has a solution if and only if $\varepsilon = 1$, or $\varepsilon = -1$ and $\mu \neq \pm 1$, or $\varepsilon = -1$ and $\mu = \pm 1$ and $z^{\mathrm{H}}\bar{r} = 0$. For $\varepsilon = 1$ or for $\varepsilon = -1$ and $\mu \neq \pm 1$, the optimal solution to* (4.7) *that minimizes $\sum_{\ell=0}^{d/2} |b_{11}^{(\ell)}|^2$ is given by*

$$b_{11}^{(d/2)} = \sigma_\varepsilon \frac{\rho_{d/2}\, \bar{\mu}^{d/2} \delta_1}{\Phi_{\text{even}}}, \tag{4.14}$$

$$b_{11}^{(\ell)} = \frac{\rho_\ell \left(\bar{\mu}^\ell + \varepsilon \bar{\mu}^{d-\ell}\right) \delta_1}{\Phi_{\text{even}}} \quad \textit{for } \ell = 0, 1, \ldots, d/2 - 1 \tag{4.15}$$

satisfying

$$\sum_{\ell=0}^{d/2} \left|b_{11}^{(\ell)}\right|^2 = \frac{|\delta_1|^2}{\Phi_{\text{even}}}. \tag{4.16}$$

For $\varepsilon = -1$ and $\mu = \pm 1$ and $z^{\mathrm{H}}\bar{r} = 0$, the optimal solution is all $b_{11}^{(\ell)} = 0$.

Proof Note all $\rho_\ell > 0$ by the assumption (2.19). For $\varepsilon = 1$, $\Phi_{\text{even}} > 0$ always; For $\varepsilon = -1$, $\Phi_{\text{even}} > 0$ if and only if $\mu \neq \pm 1$. It can be verified that $b_{ij}^{(\ell)}$'s given by (4.14) and (4.15) satisfy (4.7) whenever $\Phi_{\text{even}} > 0$. On the other hand, (4.7) implies, by the Cauchy–Schwarz inequality,

$$|\delta_1|^2 \le \Phi_{\text{even}} \left[\sum_{\ell=0}^{d/2} \left|b_{11}^{(\ell)}\right|^2\right]$$

for any solution $\{b_{ij}^{(\ell)}\}$ to (4.7).

For $\varepsilon = -1$ if $\mu = \pm 1$, then (4.7) is consistent if and only if $\delta_1 = 0$ (or equivalently $z^{\mathrm{H}}\bar{r} = 0$) for which case the optimal solution is $b_{11}^{(\ell)} = 0$ for all ℓ. □

Lemma 4.2 *Equation* (4.8) *which is for even d always has a solution. Its optimal solution that minimizes* $2|b_{12}^{(d/2)}|^2 + \sum_{\ell=0}^{d/2-1}\left(|b_{12}^{(\ell)}|^2 + |b_{21}^{(\ell)}|^2\right)$ *is given by*

$$\sqrt{2}\, b_{12}^{(d/2)} = \varepsilon \frac{\rho_{d/2}\, \bar{\mu}^{d/2}\delta_2/\sqrt{2}}{\Psi_{\text{even}}}, \tag{4.17}$$

$$b_{12}^{(\ell)} = \varepsilon \frac{\rho_\ell\, \bar{\mu}^{d-\ell}\delta_2}{\Psi_{\text{even}}}, \quad b_{21}^{(\ell)} = \frac{\rho_\ell\, \bar{\mu}^{\ell}\delta_2}{\Psi_{\text{even}}} \quad \textit{for } \ell = 0, 1, \ldots, d/2-1 \tag{4.18}$$

satisfying

$$2|b_{12}^{(d/2)}|^2 + \sum_{\ell=0}^{d/2-1}\left(|b_{12}^{(\ell)}|^2 + |b_{21}^{(\ell)}|^2\right) = \frac{|\delta_2|^2}{\Psi_{\text{even}}}. \tag{4.19}$$

Proof $\Psi_{\text{even}} > 0$ always. It can be verified that $b_{ij}^{(\ell)}$'s given by (4.17) and (4.18) satisfy (4.8). On the other hand, (4.8) implies, by the Cauchy–Schwarz inequality,

$$|\delta_2|^2 \le \Psi_{\text{even}}\left[2\left|b_{12}^{(d/2)}\right|^2 + \sum_{\ell=0}^{d/2-1}\left(\left|b_{12}^{(\ell)}\right|^2 + \left|b_{21}^{(\ell)}\right|^2\right)\right]$$

for any solution $\{b_{ij}^{(\ell)}\}$ to (4.8). □

Lemma 4.3 *Equation* (4.10) *which is for odd d has a solution if and only if either* $\mu \neq -\varepsilon$ *or* $\mu = -\varepsilon$ *and* $z^{\mathrm{H}}\bar{r} = 0$. *For* $\mu \neq -\varepsilon$, *then the optimal solution to* (4.10) *that minimizes* $\sum_{\ell=0}^{(d-1)/2}|b_{11}^{(\ell)}|^2$ *is given by*

$$b_{11}^{(\ell)} = \frac{\rho_\ell\,\left(\bar{\mu}^{\ell} + \varepsilon\bar{\mu}^{d-\ell}\right)\delta_1}{\Phi_{\text{even}}}, \quad \textit{for } \ell = 0, 1, \ldots, (d-1)/2 \tag{4.20}$$

satisfying

$$\sum_{\ell=0}^{(d-1)/2}|b_{11}^{(\ell)}|^2 = \frac{|\delta_1|^2}{\Phi_{\text{odd}}}. \tag{4.21}$$

For $\mu = -\varepsilon$ *and* $z^{\mathrm{H}}\bar{r} = 0$, *the optimal solution is all* $b_{11}^{(\ell)} = 0$.

Proof Since d is odd, $\Phi_{\text{odd}} = 0$ if and only if $\mu = -\varepsilon$. It can be verified that $b_{ij}^{(\ell)}$'s given by (4.20) satisfy (4.10) when $\Phi_{\text{even}} > 0$. On the other hand, (4.10) implies, by the Cauchy–Schwarz inequality,

$$|\delta_1|^2 \le \Phi_{\text{odd}}\left[\sum_{\ell=0}^{(d-1)/2}|b_{11}^{(\ell)}|^2\right]$$

for any solution $\{b_{ij}^{(\ell)}\}$ to (4.10).

For $\mu = -\varepsilon$, then (4.10) is consistent if and only if $\delta_1 = 0$ (or equivalently $z^{\mathrm{H}}\bar{r} = 0$) for which case the optimal solution is $b_{11}^{(\ell)} = 0$ for all ℓ. □

Lemma 4.4 *Equation* (4.11) *which is for odd d always has a solution. Its optimal solution that minimizes* $\sum_{\ell=0}^{(d-1)/2}\left(|b_{12}^{(\ell)}|^2 + |b_{21}^{(\ell)}|^2\right)$ *is given by*

$$b_{12}^{(\ell)} = \varepsilon\frac{\rho_\ell\,\bar{\mu}^{d-\ell}\delta_2}{\Psi_{\mathrm{odd}}}, \quad b_{21}^{(\ell)} = \frac{\rho_\ell\,\bar{\mu}^{\ell}\delta_2}{\Psi_{\mathrm{odd}}} \quad \text{for } \ell = 0, 1, \ldots, (d-1)/2 \tag{4.22}$$

satisfying

$$\sum_{\ell=0}^{(d-1)/2}\left(\left|b_{12}^{(\ell)}\right|^2 + \left|b_{21}^{(\ell)}\right|^2\right) = \frac{|\delta_2|^2}{\Psi_{\mathrm{odd}}}. \tag{4.23}$$

Proof $\Psi_{\mathrm{odd}} > 0$ always. It can be verified that $b_{ij}^{(\ell)}$'s given by (4.22) satisfy (4.11). On the other hand, (4.11) implies, by the Cauchy–Schwarz inequality,

$$|\delta_2|^2 \le \Psi_{\mathrm{odd}}\left[\sum_{\ell=0}^{(d-1)/2}\left(\left|b_{12}^{(\ell)}\right|^2 + \left|b_{21}^{(\ell)}\right|^2\right)\right]$$

for any solution $\{b_{ij}^{(\ell)}\}$ to (4.11). □

Remark 4.1 Our solutions to (4.7) and (4.10) by Lemmas 4.1 and 4.3 are all that are needed for dealing with the scalar polynomial case, i.e., when $n = 1$.

5 Numerical examples

In this section we test some numerical examples to illustrate our structured backward error compared with the existing (unstructured) backward errors. For convenience, we consider PQEP only: $Q(\lambda) \equiv \varepsilon\lambda^2 A_0^{\star} + \lambda A_1 + A_0$ with also $A_1^{\star} = \varepsilon A_1$, and set $\rho_i = \|A_i\|_2$.

By the existing (unstructured) backward error, we mean the one from Tisseur [18] who defined the (unstructured) backward error of an approximate eigenpair $\{\mu, z\}$ of $Q(\lambda)$ as

$$\begin{aligned}\eta(\mu, z) = \min\{\delta\ :\ &[Q(\mu) + \Delta Q(\mu)]z = 0,\\ &||\Delta A_2||_2 \le \delta\rho_0,\ ||\Delta A_1||_2 \le \delta\rho_1,\ ||\Delta A_0||_2 \le \delta\rho_0\},\end{aligned}$$

where $\Delta Q(\lambda) = \lambda^2(\Delta A_2) + \lambda(\Delta A_1) + \Delta A_0$ with no constraints imposed on ΔA_i. An explicit expression for $\eta(\mu, z)$ in terms of $r = -Q(\mu)z$ is given by [18]

$$\eta(\mu, z) = \frac{||r||_2}{(|\mu|^2\rho_0 + |\mu|\rho_1 + \rho_0)\,||z||_2}. \tag{5.1}$$

Take $n = 10$ and choose $A_0, A_1 \in \mathbb{C}^{10\times 10}$ randomly as by this piece of MATLAB code for $\star = \mathrm{H}$:

```
A0 = randn(n)+i*randn(n); A2 = varepsilon * A0';
A1 = randn(n)+i*randn(n); A1 = A1+varepsilon*A1';
```

For $\star = \mathrm{T}$, all needed is to replace `A0'` and `A1'` by `A0.'` and `A1.'`, respectively. We compute $2n$ approximate eigenpairs $\{\mu, z\}$ of $Q(\lambda)$ by MATLAB's `eig(···)` after linearizing $Q(\lambda)$. With $\{\mu, z\}$ running through all $2n$ approximate eigenpairs, we plots various quantities of interest in Figs. 1 and 2.

Figure 1 is for $\star = \mathrm{H}$. It is worth pointing out that in our many numerical runs with different random matrices constructed as such (and different n as well), there are always some eigenvalues with $|\mu|$ being 1 in the working (IEEE double) precision. As our analysis in Sect. 3 shows, if $|\mu| = 1$, the matrix Z is singular and there may not exist a structured backward error unless $z^{\mathrm{H}}r/(\sqrt{\varepsilon}\mu^{d/2})$ is real. But this condition is nearly impossible to verify when the approximate eigenpair is so accurate in the working precision that r may have no correct significant digits. So in the numerical results presented in Fig. 1, we purposely perturb the computed μ relatively by 10^{-8} (i.e., perturb μ to $\mu(1 + 10^{-8})$) to make $|\mu| - 1$ away from 0 by no less than about 10^{-8}. Doing so, however, does not prevent us from making our point with the numerical results, namely Δ_{F} is very sensitive with respect to $||\mu| - 1|$ being near 0 or not. Figure 1 clearly shows when $||\mu| - 1|$ is away from 0, Δ_{F} is not much worse than $\eta(\mu, z)$, and it appears that Δ_{F} is proportional to the reciprocal of $||\mu| - 1|$.

Figure 2 is for $\star = \mathrm{T}$. We did not encounter the critical cases $\mu = \pm 1$ as revealed in our analysis in Sect. 4 during our many numerical runs. The numerical results presented in this figure do not include any critical case, unfortunately. Because of the lack of critical cases, Δ_{F} is not much worse than $\eta(\mu, z)$ here.

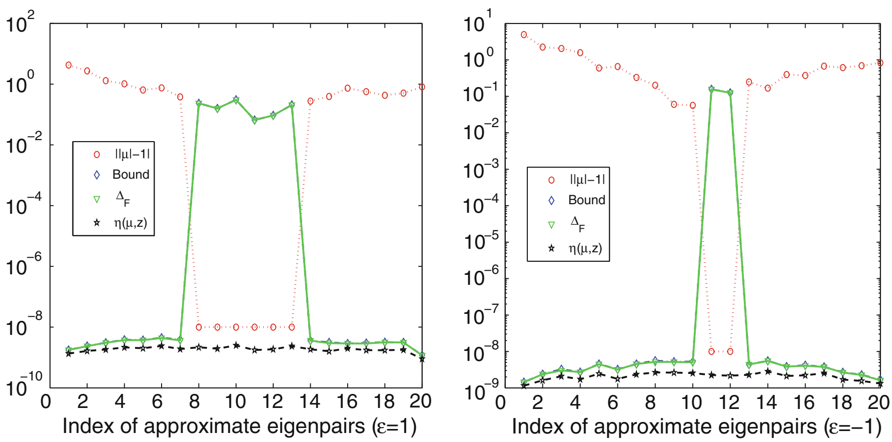

Fig. 1 Backward errors for the case $\star = \mathrm{H}$; *Left:* $\varepsilon = 1$ and *Right:* $\varepsilon = -1$. Plotted quantities are $||\mu| - 1|$, bounds by (3.11), Δ_{F} by combining (3.24) and (3.42), and Tisseur's $\eta(\mu, z)$ in (5.1). Bounds by (3.11) and Δ_{F} are visually indistinguishable from the plotted lines. Here, all computed μ by `eig(···)` are perturbed relatively by 10^{-8} to make sure that the computed residuals have several correct significant digits

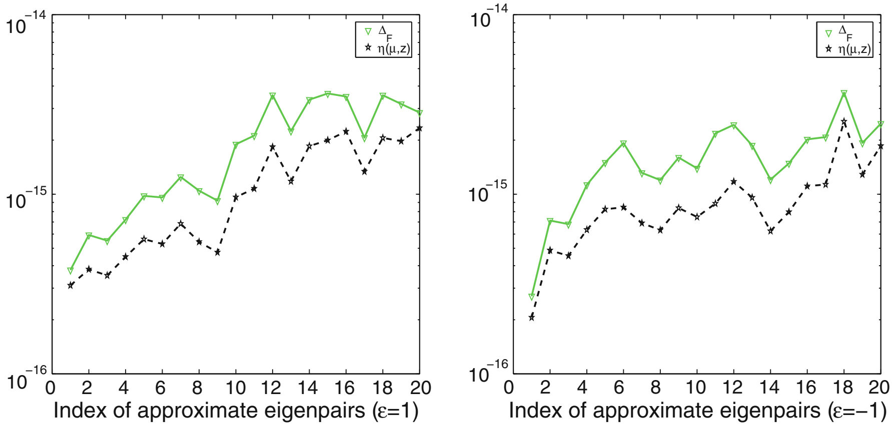

Fig. 2 Backward errors for the case $\star = \top$; *Left:* $\varepsilon = 1$ and *Right:* $\varepsilon = -1$. Plotted quantities are Δ_F by (4.5), and Tisseur's $\eta(\mu, z)$ in (5.1)

6 Concluding remarks

We have presented a detailed structured backward error analysis for an approximate eigenpair $\{\mu, z\}$ of PPEP (1.4) by solving problem (1.5). Computable formulas and bounds for the structured backward errors are obtained. These formulas and bounds show distinctive features of PPEP from the usual PEP as investigated by Tisseur [18] and Liu and Wang [13], namely there are critical cases ($|\mu| = 1$ for $\star = \mathrm{H}$ and $\mu = \pm 1$ for $\star = \top$) for which structured backward errors may not exist. When structured backward errors do exist, in the worst case the structured backward error is inversely proportional to the square root of, for $\star = \mathrm{H}$,

$$\sum_{\ell=0}^{d/2-1} \rho_\ell^2 \left(|\mu|^\ell - |\mu|^{d-\ell}\right)^2 \quad \text{for even } d, \text{ and}$$

$$\sum_{\ell=0}^{(d-1)/2} \rho_\ell^2 \left(|\mu|^\ell - |\mu|^{d-\ell}\right)^2 \quad \text{for odd } d,$$

and for $\star = \top$,

$$\sum_{\ell=0}^{d/2-1} \rho_\ell^2 \left|\mu^\ell + \varepsilon\mu^{d-\ell}\right|^2 + \sigma_\varepsilon \rho_{d/2}^2 |\mu|^d \quad \text{for even } d, \text{ and}$$

$$\sum_{\ell=0}^{(d-1)/2} \rho_\ell^2 \left|\mu^\ell + \varepsilon\mu^{d-\ell}\right|^2 \quad \text{for odd } d.$$

The first two vanish at $|\mu| = 1$, while the last two possibly vanish at $\mu = \pm 1$.

In the PEP case, Tisseur [18] asked and solved what the backward error for an approximated eigenvalue is by minimizing the backward errors for an approximate eigenpair $\{\mu, z\}$ over all possible vectors z. We also attempted to solve a similar question, namely what the *structured* backward error for an approximated eigenvalue in the PPEP case is. But it appears that our question is not as simple as in the PEP case because our formulas for structured backward errors for PPEP depend on the approximate eigenvector z in a more complicated way than Tisseur's. Also in the PEP case, Tisseur [18] asked and solved what the backward errors for an approximate eigentriplet (an eigenvalue and its associated left and right eigenvectors). We attempted to solve a similar question, too, but did not succeed.

Acknowledgments Li is supported in part by NSF grants DMS-0702335 and DMS-0810506 and by the National Center for Theoretical Sciences at National Tsing-Hua University, Taiwan, where this work was initiated while he was visiting the center in June 2008.

References

1. Bora, S.: Structured eigenvalue condition number and backward error of a class of polynomial eigenvalue problems. SIAM J. Matrix Anal. Appl. **31**(3), 900–917 (2009)
2. Bunch, J.R., Demmel, J., Van Loan, C.F.: The strong stability of algorithms for solving symmetric linear systems. SIAM J. Matrix Anal. Appl. **10**(4), 494–499 (1989)
3. Byers, R., Mackey, D.S., Mehrmann, V., Xu, H.: Symplectic, BVD, and palindromic approaches to discrete-time control problems. Technical report, Preprint 14-2008, Institute of Mathematics, Technische Universität Berlin, 2008
4. Chu, E.K.-W., Hwang, T.-M., Lin, W.-W., Wu, C.-T.: Vibration of fast trains, palindromic eigenvalue problems and structure-preserving doubling algorithms. J. Comput. Appl. Math. **229**, 237–252 (2008)
5. Higham, D.J., Higham, N.J.: Backward error and condition of structured linear systems. SIAM J. Matrix Anal. Appl. **13**(1), 162–175 (1992)
6. Higham, D.J., Higham, N.J.: Structured backward error and condition of generalized eigenvalue problems. SIAM J. Matrix Anal. Appl. **20**(2), 493–512 (1998)
7. Higham, N.J., Tisseur, F., Van Dooren, P.: Detecting a definite Hermitian pair and a hyperbolic or elliptic quadratic eigenvalue problem, and associated nearness problems. Linear Algebra Appl. **351–352**, 455–474 (2002)
8. Hilliges, A.: Numerische Lösung von quadratischen Eigenwertproblemen mit Anwendungen in der Schiendynamik. Master's thesis, Technical University Berlin, Germany, July 2004
9. Hilliges, A., Mehl, C., Mehrmann, V.: On the solution of palindramic eigenvalue problems. In: Proceedings of 4th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS), Jyväskylä, Finland, 2004
10. Hochstenbach, M.E., Plestenjak, B.: Backward error, condition numbers, and pseudospectra for the multiparameter eigenvalue problem. Linear Algebra Appl. **375**, 63–81 (2003)
11. Huang, T.-M., Lin, W.-W., Qian, J.: Numerically stable, structure-preserving algorithms for palindromic quadratic eigenvalue problems arising from vibration of fast trains. SIAM J. Matrix Anal. Appl. **30**(4), 1566–1592 (2009)
12. Ipsen, I.C.F.: Accurate eigenvalues for fast trains. SIAM News **37**, 1–2 (2004)
13. Liu, X.-G., Wang, Z.-X.: A note on the backward errors for Hermite eigenvalue problems. Appl. Math. Comput. **165**(2), 405–417 (2005)
14. Mackey, D.S., Mackey, N., Mehl, C., Mehrmann, V.: Structured polynomial eigenvalue problems: good vibrations from good linearizations. SIAM J. Matrix Anal. Appl. **28**, 1029–1051 (2006)
15. Mackey, D.S., Mackey, N., Mehl, C., Mehrmann, V.: Vector spaces of linearizations for matrix polynomials. SIAM J. Matrix Anal. Appl. **28**, 971–1004 (2006)
16. Schröder, C.: A QR-like algorithm for the palindromic eigenvalue problem. Technical report, Preprint 388, TU Berlin, Matheon, Germany, 2007

17. Schröder, C.: URV decomposition based structured methods for palindromic and even eigenvalue problems. Technical report, Preprint 375, TU Berlin, MATHEON, Germany, 2007
18. Tisseur, F.: Backward error and condition of polynomial eigenvalue problems. Linear Algebra Appl. **309**, 339–361 (2000)
19. Tisseur, F.: Stability of structured Hamiltonian eigensolvers. SIAM J. Matrix Anal. Appl. **23**(1), 103–125 (2001)
20. Tisseur, F., Meerbergen, K.: The quadratic eigenvalue problem. SIAM Rev. **43**, 235–386 (2001)
21. Xu, H.: On equivalence of pencils from discrete-time and continuous-time control. Linear Algebra Appl. **414**, 97–124 (2006)
22. Zaglmayr, S.: Eigenvalue problems in saw-filter simulations. Diplomarbeit, Institute of Computational Mathematics, Johannes Kepler University Linz, Linz, Austria, 2002