

***Backward Error of Polynomial Eigenproblems
Solved by Linearization***

Higham, Nicholas J. and Li, Ren-Cang and Tisseur,
Françoise

2007

MIMS EPrint: **2006.137**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

BACKWARD ERROR OF POLYNOMIAL EIGENPROBLEMS SOLVED BY LINEARIZATION*

NICHOLAS J. HIGHAM[†], REN-CANG LI[‡], AND FRANÇOISE TISSEUR[†]

Abstract. The most widely used approach for solving the polynomial eigenvalue problem $P(\lambda)x = (\sum_{i=0}^m \lambda^i A_i)x = 0$ in $n \times n$ matrices A_i is to linearize to produce a larger order pencil $L(\lambda) = \lambda X + Y$, whose eigensystem is then found by any method for generalized eigenproblems. For a given polynomial P , infinitely many linearizations L exist and approximate eigenpairs of P computed via linearization can have widely varying backward errors. We show that if a certain one-sided factorization relating L to P can be found then a simple formula permits recovery of right eigenvectors of P from those of L , and the backward error of an approximate eigenpair of P can be bounded in terms of the backward error for the corresponding approximate eigenpair of L . A similar factorization has the same implications for left eigenvectors. We use this technique to derive backward error bounds depending only on the norms of the A_i for the companion pencils and for the vector space $\mathbb{DL}(P)$ of pencils recently identified by Mackey, Mackey, Mehl, and Mehrmann. In all cases, sufficient conditions are identified for an optimal backward error for P . These results are shown to be entirely consistent with those of Higham, Mackey, and Tisseur on the conditioning of linearizations of P . Other contributions of this work are a block scaling of the companion pencils that yields improved backward error bounds; a demonstration that the bounds are applicable to certain structured linearizations of structured polynomials; and backward error bounds specialized to the quadratic case, including analysis of the benefits of a scaling recently proposed by Fan, Lin, and Van Dooren. The results herein make no assumptions on the stability of the method applied to L or whether the method is direct or iterative.

Key words. backward error, scaling, eigenvector, matrix polynomial, matrix pencil, linearization, companion form, quadratic eigenvalue problem, alternating, palindromic

AMS subject classifications. 65F15, 15A18

DOI. 10.1137/060663738

1. Introduction. The polynomial eigenvalue problem (PEP) is to find scalars λ and nonzero vectors x and y satisfying $P(\lambda)x = 0$ and $y^*P(\lambda) = 0$, where

$$(1.1) \quad P(\lambda) = \sum_{i=0}^m \lambda^i A_i, \quad A_i \in \mathbb{C}^{n \times n}, \quad A_m \neq 0$$

is a matrix polynomial of degree m . Here, x and y are right and left eigenvectors corresponding to the eigenvalue λ . We will assume throughout that P is regular, that is, $\det P(\lambda) \not\equiv 0$.

The standard way of solving this problem is to convert P into a linear polynomial

$$L(\lambda) = \lambda X + Y, \quad X, Y \in \mathbb{C}^{mn \times mn}$$

*Received by the editors June 26, 2006; accepted for publication (in revised form) by P. Van Dooren June 7, 2007; published electronically December 7, 2007.

<http://www.siam.org/journals/simax/29-4/66373.html>

[†]School of Mathematics, The University of Manchester, Manchester, M13 9PL, UK (higham@ma.man.ac.uk, <http://www.ma.man.ac.uk/~higham/>, ftisseur@ma.man.ac.uk, <http://www.ma.man.ac.uk/~ftisseur/>). The work of the first author was supported by a Royal Society-Wolfson Research Merit Award. The work of the first and third authors was supported by Engineering and Physical Sciences Research Council grants GR/S31693 and EP/D079403.

[‡]Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019 (rccli@uta.edu). The work of this author was supported in part by the National Science Foundation CAREER award under grant CCR-9875201 and by the National Science Foundation under grant DMS-0510664.

with the same spectrum as P and solve the eigenproblem for L . This generalized eigenproblem is usually solved with the QZ algorithm [20] for small to medium size problems or a projection method for large sparse problems [1]. That L has the same spectrum as P is assured if

$$(1.2) \quad E(\lambda)L(\lambda)F(\lambda) = \begin{bmatrix} P(\lambda) & 0 \\ 0 & I_{(m-1)n} \end{bmatrix}$$

for some unimodular $E(\lambda)$ and $F(\lambda)$. (A matrix polynomial $E(\lambda)$ is unimodular if its determinant is a nonzero constant, independent of λ .) Such an L is called a *linearization* of $P(\lambda)$ [5, sec. 7.2]. As an example, the pencil

$$(1.3) \quad C_1(\lambda) = \lambda \begin{bmatrix} A_3 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} + \begin{bmatrix} A_2 & A_1 & A_0 \\ -I & 0 & 0 \\ 0 & -I & 0 \end{bmatrix}$$

can be shown to be a linearization for the cubic $P(\lambda) = \lambda^3 A_3 + \lambda^2 A_2 + \lambda A_1 + A_0$; it is known as the *first companion* linearization.

Among the infinitely many linearizations L of P we are interested in those whose right and left eigenvectors permit easy recovery of the corresponding eigenvectors of P . For example, if (x, y) and (z, w) denote pairs of right and left eigenvectors of the cubic $P(\lambda)$ and its companion linearization $C_1(\lambda)$, respectively, associated with the simple, finite eigenvalue λ , then

$$(1.4) \quad (z, w) = \left(\begin{bmatrix} \lambda^2 x \\ \lambda x \\ x \end{bmatrix}, \begin{bmatrix} y \\ (\bar{\lambda}A_3^* + A_2^*)y \\ (\bar{\lambda}^2 A_3^* + \bar{\lambda}A_2^* + A_1^*)y \end{bmatrix} \right),$$

so that x can be recovered from one of the first two blocks (if $\lambda \neq 0$) or the third block of n components of z , and y can be recovered from the first n components of w . This correspondence extends to all eigenvalues and arbitrary m , as we will explain in section 3.

In practice, the eigenpairs of L are not computed exactly because of rounding errors and, in the case of iterative methods, truncation errors. For a given approximate eigenpair of L , it is important to know how good an approximate eigenpair of P will be produced. Here, “good” can have various meanings; in particular, it can refer to the relative error of the eigenvalue or the backward error of the eigenpair. The relative error question has been investigated by Higham, Mackey, and Tisseur [7], by analyzing the conditioning of both the polynomial P and the linearization L . The purpose of the present work is to investigate the backward error for a wide variety of linearizations. Two key aspects of this task can be seen by considering the companion pencil (1.3). First, a small but arbitrary perturbation to C_1 , such as that introduced by the QZ algorithm, does not respect the zero and identity blocks and so may not correspond to a small perturbation of P . Second, the block from which the approximate eigenvector is recovered will influence the backward error.

Our work builds on that of Tisseur [22], who shows that solving a quadratic eigenvalue problem (QEP) by applying a numerically stable method to the companion linearization can be backward unstable, but that stability is guaranteed if all the coefficient matrices have unit norm.

In section 2.1 we define the backward error η_P of an approximate eigenpair and eigentriple of P for the polynomial both in the λ -form (1.1) and in homogeneous (α, β) -form. In section 2.2 we show that given appropriate one-sided factorizations relating

a linearization L to the original polynomial P , we can bound the backward error of an approximate eigenpair of P in terms of the backward error of the approximate eigenpair of L from which it was obtained. The bounds have the useful feature of separating the dependence on L , P , and (α, β) from the dependence on how the (right or left) eigenvector is recovered.

In section 3 we introduce the first and second companion linearizations and the vector spaces \mathbb{L}_1 and \mathbb{L}_2 of pencils associated with P . As a by-product of our analysis we obtain in section 3.1 new formulae for recovering a left (right) eigenvector of P from one of a linearization in \mathbb{L}_1 (\mathbb{L}_2). In section 3.2 we obtain backward error bounds for the companion pencils and deduce sufficient conditions for a small backward error η_P . We show in section 3.3 that applying a block scaling to the companion pencils yields smaller backward error bounds when $\max_i \|A_i\|_2$ is much different from 1. The vector space $\mathbb{DL}(P) = \mathbb{L}_1(P) \cap \mathbb{L}_2(P)$ is then considered in section 3.4, where bounds of the same form as for the block-scaled companion pencils are obtained. In section 3.5 we explain how the backward error results provide essentially the same guidance on optimal choice of linearizations as the condition number bounds of Higham, Mackey, and Tisseur [7]. In section 4 we show that the results of section 3 also apply to certain structured linearizations of structured polynomials.

The special case of quadratic polynomials $\lambda^2 A + \lambda B + C$ is studied in detail in section 5, concentrating on the companion linearization and the $\mathbb{DL}(P)$ basis pencils L_1 and L_m . Bounds for η_P are obtained and then specialized to exploit a scaling procedure recently proposed by Fan, Lin, and Van Dooren [2]. The bounds involve a growth factor ω that is shown to be bounded by $1 + \tau$, where $\tau = \|B\|_2 / \sqrt{\|A\|_2 \|C\|_2}$. Our analysis improves upon that in [2], which contains a growth term $\max(1 + \tau, 1 + \tau^{-1})$. The bounds are particularly satisfactory for elliptic QEPs and, more generally, QEPs that are not too heavily damped. Numerical experiments illustrating these and other aspects of the theory are given in section 6.

Finally, we note that our results are of interest even in the case $n = 1$, although we will not consider this case specifically here. The roots of a scalar polynomial, p , are often found by computing the eigenvalues of a corresponding companion matrix, C . Our analysis provides new bounds on the backward errors of the computed roots of p in terms of the backward errors of the computed eigenvalues of C .

2. Backward errors.

2.1. Definition and notation. The normwise backward error of an approximate (right) eigenpair (x, λ) of $P(\lambda)$, where λ is finite, is defined by

$$(2.1) \quad \eta_P(x, \lambda) = \min\{\epsilon : (P(\lambda) + \Delta P(\lambda))x = 0, \|\Delta A_i\|_2 \leq \epsilon \|A_i\|_2, i = 0:m\},$$

where $\Delta P(\lambda) = \sum_{i=0}^m \lambda^i \Delta A_i$. Tisseur [22, Thm. 1] obtained the explicit formula

$$(2.2) \quad \eta_P(x, \lambda) = \frac{\|P(\lambda)x\|_2}{(\sum_{i=0}^m |\lambda^i| \|A_i\|_2) \|x\|_2}.$$

Similarly, for an approximate left eigenpair (y^*, λ) , we have

$$(2.3) \quad \eta_P(y^*, \lambda) := \min\{\epsilon : y^*(P(\lambda) + \Delta P(\lambda)) = 0, \|\Delta A_i\|_2 \leq \epsilon \|A_i\|_2, i = 0:m\}$$

$$(2.4) \quad = \frac{\|y^* P(\lambda)\|_2}{(\sum_{i=0}^m |\lambda^i| \|A_i\|_2) \|y\|_2}.$$

Also of interest is the backward error of the approximate triplet (x, y, λ) [22, Thm. 4]:

$$(2.5) \quad \eta_P(x, y^*, \lambda) := \min\{ \epsilon : (P(\lambda) + \Delta P(\lambda))x = 0, y^*(P(\lambda) + \Delta P(\lambda)) = 0,$$

$$\|\Delta A_i\|_2 \leq \epsilon \|A_i\|_2, i = 0:m \}$$

$$(2.6) \quad = \max(\eta_P(x, \lambda), \eta_P(y^*, \lambda)).$$

We make two comments on notation. As an argument of η , a left eigenvector is written as a row vector to distinguish it from a right eigenvector. Symbols such as λ, x , and y will denote both exact and (more often) approximate quantities, with the context making clear which usage is in effect. The alternative of using a tilde to denote approximate quantities leads to rather cumbersome formulae.

In order to define backward errors valid for all λ , including ∞ , we rewrite the polynomial in the homogeneous form

$$P(\alpha, \beta) = \sum_{i=0}^m \alpha^i \beta^{m-i} A_i$$

and identify λ with any pair $(\alpha, \beta) \neq (0, 0)$ for which $\lambda = \alpha/\beta$. The definitions (2.1), (2.3), and (2.5) are trivially rewritten in terms of α and β . Using $P(\alpha, \beta) = \beta^m P(\alpha/\beta)$ for $\beta \neq 0$, we find that in place of (2.2), (2.4), and (2.6) we have

$$(2.7) \quad \eta_P(x, \alpha, \beta) = \frac{\|P(\alpha, \beta)x\|_2}{(\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2) \|x\|_2},$$

$$(2.8) \quad \eta_P(y^*, \alpha, \beta) = \frac{\|y^* P(\alpha, \beta)\|_2}{(\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2) \|y\|_2},$$

$$(2.9) \quad \eta_P(x, y^*, \alpha, \beta) = \max(\eta_P(x, \alpha, \beta), \eta_P(y^*, \alpha, \beta)).$$

Note that these expressions are independent of the choice of α and β representing the eigenvalue; that is, a scaling $\alpha \leftarrow \theta\alpha, \beta \leftarrow \theta\beta$ with $\theta \neq 0$ leaves the expressions unchanged.

2.2. Bounding the backward error for P relative to that for L . Let $L(\lambda) = \lambda X + Y$ be a linearization of $P(\lambda)$. For approximate right eigenvectors z of L and x of P , both corresponding to an approximate eigenvalue (α, β) , our aim is to compare $\eta_P(x, \alpha, \beta)$ with

$$(2.10) \quad \eta_L(z, \alpha, \beta) = \frac{\|L(\alpha, \beta)z\|_2}{(|\alpha|\|X\|_2 + |\beta|\|Y\|_2)\|z\|_2},$$

which is obtained by applying (2.7) to $L(\alpha, \beta) = \alpha X + \beta Y$. Of course, this comparison is possible only if there is some well-defined relation between x and z . Such a relation, and a means for bounding η_P , both follow from one key assumption: that we can find an $n \times nm$ matrix polynomial $G(\alpha, \beta)$ such that

$$(2.11) \quad G(\alpha, \beta)L(\alpha, \beta) = g^T \otimes P(\alpha, \beta)$$

for some nonzero $g \in \mathbb{C}^m$, where \otimes denotes the Kronecker product [15, sec. 12.1]. Necessarily, $G(\alpha, \beta)$ will have degree $m - 1$. Note that this is a one-sided transformation as opposed to the two-sided transformation in the definition of linearization. Then we have

$$(2.12) \quad G(\alpha, \beta)L(\alpha, \beta)z = (g^T \otimes P(\alpha, \beta))z = P(\alpha, \beta)(g^T \otimes I_n)z,$$

where the latter equation relies on g^T being a row vector. Thus if z is an eigenvector of L then

$$(2.13) \quad x := (g^T \otimes I_n)z = \sum_{i=1}^m g_i z_i, \quad z_i := z((i-1)n + 1:in)$$

is an eigenvector of P , provided that x is nonzero. This latter requirement is not satisfied in general but will be proved for some important classes of linearizations. As an example, for the first companion linearization $C_1(\alpha, \beta) = \beta^3 C_1(\alpha/\beta)$ in (1.3), it is easily checked that $G(\alpha, \beta) = [\alpha^2 I \quad -(\beta^2 A_0 + \alpha\beta A_1) \quad -\alpha\beta A_0]$ satisfies (2.11) with $g = e_1$, the first column of the identity matrix, and that if z is a right eigenvector of C_1 and $\alpha \neq 0$ then $x = z_1 = z(1:n) \neq 0$ is a right eigenvector for P (cf. (1.4)).

Suppose now that (2.11) is satisfied, an approximate right eigenvector z of L is given, and x is given by (2.13). Then, by (2.7), (2.10), and (2.12),

$$(2.14) \quad \eta_P(x, \alpha, \beta) \leq \frac{\|G(\alpha, \beta)\|_2 \|L(\alpha, \beta)z\|_2}{(\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2) \|x\|_2} \\ \leq \frac{|\alpha| \|X\|_2 + |\beta| \|Y\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2} \cdot \frac{\|G(\alpha, \beta)\|_2 \|z\|_2}{\|x\|_2} \cdot \eta_L(z, \alpha, \beta).$$

This bound largely separates the dependence on L , P , and (α, β) (in the first term) from the dependence on G and z (in the second term).

For left eigenvectors the appropriate analogue of the assumption (2.11) is that there exists an $mn \times n$ matrix polynomial $H(\alpha, \beta)$ such that

$$(2.15) \quad L(\alpha, \beta)H(\alpha, \beta) = h \otimes P(\alpha, \beta)$$

for some nonzero $h \in \mathbb{C}^m$. We then have, for $w \in \mathbb{C}^{mn}$,

$$(2.16) \quad w^* L(\alpha, \beta)H(\alpha, \beta) = w^*(h \otimes P(\alpha, \beta)) = w^*(h \otimes I_n)P(\alpha, \beta).$$

Hence if w is a left eigenvector of L then

$$(2.17) \quad y := (h^* \otimes I_n)w = \sum_{i=1}^m \bar{h}_i w_i, \quad w_i := w((i-1)n + 1:in)$$

is a left eigenvector of P , provided that it is nonzero. From (2.8) and (2.17) we obtain for an approximate left eigenvector w of L the bound

$$(2.18) \quad \eta_P(y^*, \alpha, \beta) \leq \frac{|\alpha| \|X\|_2 + |\beta| \|Y\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2} \cdot \frac{\|H(\alpha, \beta)\|_2 \|w\|_2}{\|y\|_2} \cdot \eta_L(w^*, \alpha, \beta).$$

In the rest of this paper we show that one or both of assumptions (2.11) and (2.15) are satisfied for a wide class of linearizations, and we study the upper bounds (2.14) and (2.18).

3. Unstructured linearizations. We first concentrate on general, unstructured matrix polynomials, treating companion and $\mathbb{DL}(P)$ linearizations.

Associated with P are two companion pencils, $C_1(\lambda) = \lambda X_1 + Y_1$ and $C_2(\lambda) = \lambda X_2 + Y_2$, called the first and second companion forms [15, sec. 14.1], respectively,

where

$$\begin{aligned}
 X_1 = X_2 &= \text{diag}(A_m, I_n, \dots, I_n), \\
 (3.1) \quad Y_1 &= \begin{bmatrix} A_{m-1} & A_{m-2} & \dots & A_0 \\ -I_n & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -I_n & 0 \end{bmatrix}, \quad Y_2 = \begin{bmatrix} A_{m-1} & -I_n & \dots & 0 \\ A_{m-2} & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & -I_n \\ A_0 & 0 & \dots & 0 \end{bmatrix}.
 \end{aligned}$$

They are widely used in practice. For example, the MATLAB function `polyeig` that solves the PEP uses the reversed first companion linearization $\text{rev}C_1(\lambda)$ of the reversed matrix polynomial $\text{rev}P(\lambda)$. The reversal operator is defined for P in (1.1) by

$$(3.2) \quad \text{rev}P(\lambda) = \lambda^m P(1/\lambda) = \sum_{i=0}^m \lambda^i A_{m-i}.$$

The companion forms have the important property that they are always linearizations [19, sec. 4].

$C_1(\lambda)$ and $C_2(\lambda)$ belong to large sets of potential linearizations recently identified by Mackey et al. [19] and studied in [6] and [19]. With the notation $\Lambda = [\lambda^{m-1}, \lambda^{m-2}, \dots, 1]^T$, these sets are

$$(3.3) \quad \mathbb{L}_1(P) = \{ L(\lambda) : L(\lambda)(\Lambda \otimes I_n) = v \otimes P(\lambda), v \in \mathbb{C}^m \},$$

$$(3.4) \quad \mathbb{L}_2(P) = \{ L(\lambda) : (A^T \otimes I_n)L(\lambda) = \tilde{v}^T \otimes P(\lambda), \tilde{v} \in \mathbb{C}^m \}.$$

There are many $L(\lambda) \in \mathbb{L}_1(P)$ corresponding to a given v , and likewise for $\mathbb{L}_2(P)$; indeed, $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ both have dimension $m(m-1)n^2 + m$ [19, Cor. 3.6]. It is easy to check that $C_1(\lambda)$ and $C_2(\lambda)$ belong to $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$, respectively, with $\tilde{v} = v = e_1$; so the pencils in \mathbb{L}_1 and \mathbb{L}_2 can be thought of as generalizations of the first and second companion forms. It is proved in [19, Prop. 3.2, Prop. 3.12, Thm. 4.7] that $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ are vector spaces and that almost all pencils in these spaces are linearizations of P .

One of the underlying reasons for the interest in \mathbb{L}_1 and \mathbb{L}_2 is that eigenvectors of P can be directly recovered from eigenvectors of linearizations in \mathbb{L}_1 and \mathbb{L}_2 . As with the backward errors, it is more convenient to use the (α, β) notation, so we define

$$\Lambda_{\alpha, \beta} = [\alpha^{m-1}, \alpha^{m-2}\beta, \dots, \beta^{m-1}]^T = \beta^{m-1} \Lambda.$$

THEOREM 3.1 (eigenvector recovery from \mathbb{L}_1 and \mathbb{L}_2).

- If $L \in \mathbb{L}_1(P)$ is a linearization of P then every right eigenvector of L with eigenvalue (α, β) is of the form $\Lambda_{\alpha, \beta} \otimes x$ for some right eigenvector x of P .
- If $L \in \mathbb{L}_2(P)$ is a linearization of P then every left eigenvector of L with eigenvalue (α, β) is of the form $\bar{\Lambda}_{\alpha, \beta} \otimes y$ for some left eigenvector y of P .

Proof. See [19, Thms. 3.8, 3.14, 4.4]. \square

Theorem 3.1 shows that from any right eigenvector z of $L \in \mathbb{L}_1$ we can read off a right eigenvector of P by looking at any nonzero subvector $z_i = z((i-1)n + 1 : in)$, and similarly a left eigenvector of $L \in \mathbb{L}_2$ yields a left eigenvector of P .

3.1. $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. It is immediate from (3.3) and (3.4) that the pencils in \mathbb{L}_1 satisfy (2.15), while those in \mathbb{L}_2 satisfy (2.11). Therefore our backward error bounds are applicable to left eigenvectors of pencils in \mathbb{L}_1 and right eigenvectors of

pencils in \mathbb{L}_2 , provided that the vectors x in (2.13) and y in (2.17) are nonzero when z and w are exact eigenvectors. In fact, for \mathbb{L}_1 and \mathbb{L}_2 (2.13) and (2.17) define a bijection between eigenvectors of the pencil and of P and so allow recovery of all the eigenvectors. The following two new results supplement the existing eigenvector recovery formulae in Theorem 3.1.

THEOREM 3.2 (left eigenvector recovery from \mathbb{L}_1). *Let $L \in \mathbb{L}_1(P)$ be a linearization of P , with vector v (necessarily nonzero) in (3.3). If w is a left eigenvector of L with eigenvalue (α, β) then*

$$(3.5) \quad y = (v^* \otimes I_n)w$$

is a left eigenvector of P with eigenvalue (α, β) . Moreover, any left eigenvector of P corresponding to (α, β) can be recovered from one of L from the formula (3.5).

Proof. Assume, first, that $\mu \equiv (\alpha, \beta)$ is finite. For arbitrary λ , premultiplying the condition defining \mathbb{L}_1 by w^* (or simply using the λ -analogue of (2.16)) gives

$$w^*L(\lambda)(A \otimes I_n) = w^*(v \otimes P(\lambda)) = w^*(v \otimes I_n)P(\lambda) =: y^*P(\lambda).$$

Since $w^*L(\mu) = 0$, it follows that $y^*P(\mu) = 0$. We therefore just have to show that $y \neq 0$. We suppose that $y = 0$ and will obtain a contradiction. If $y = 0$ then $w^*L(\lambda)(A \otimes I_n) \equiv 0$. Since L is linear, we can write

$$w^*L(\lambda) = [b_1(\lambda), b_2(\lambda), \dots, b_m(\lambda)],$$

where $b_i(\lambda) = c_i\lambda + d_i \in \mathbb{C}^{1 \times n}$ is linear. Then

$$\begin{aligned} 0 \equiv w^*L(\lambda)(A \otimes I_n) &= [b_1(\lambda), b_2(\lambda), \dots, b_m(\lambda)] \begin{bmatrix} \lambda^{m-1}I_n \\ \lambda^{m-2}I_n \\ \vdots \\ I_n \end{bmatrix} \\ &= \lambda^{m-1}b_1(\lambda) + \lambda^{m-2}b_2(\lambda) + \dots + b_m(\lambda) \\ &= \lambda^m c_1 + \lambda^{m-1}(d_1 + c_2) + \dots + \lambda(d_{m-1} + c_m) + d_m. \end{aligned}$$

Hence $c_1 = 0, d_1 = -c_2, \dots, d_{m-1} = -c_m, d_m = 0$. Then

$$0 = w^*L(\mu) = [b_1(\mu), b_2(\mu), \dots, b_m(\mu)] = [-c_2, \mu c_2 - c_3, \dots, \mu c_{m-1} - c_m, \mu c_m],$$

which implies $c_2 = c_3 = \dots = c_m = 0$. Hence $b_i(\lambda) \equiv 0$ for all i . Thus $w^*L(\lambda) \equiv 0$, which means that L is a nonregular polynomial. But by [19, Thm. 4.3], $L \in \mathbb{L}_1(P)$ being nonregular implies that L is not a linearization of P . This is a contradiction, and so $y \neq 0$, as required.

The case $\mu = \infty$ can be handled by expressing L and P in homogeneous (α, β) -form and using $\mu \equiv (1, 0)$. The details are a minor variation on those above.

Finally, consider the map $w \mapsto (v^* \otimes I_n)w$ from $\mathcal{K}_1 = \text{left ker } L(\alpha, \beta)$ to $\mathcal{K}_2 = \text{left ker } P(\alpha, \beta)$, where left ker denotes the left kernel. The first part showed that this map has kernel $\{0\}$. Since $L \in \mathbb{L}_1(P)$ is a linearization and P is regular, L is a strong linearization¹ [19, Thm. 4.3]. Hence the geometric multiplicity of any eigenvalue (including ∞) is the same for L and P [14]; that is, \mathcal{K}_1 and \mathcal{K}_2 have the same dimension. It follows that the map is a bijection, and the result is proved. \square

¹ L is a strong linearization of P if it is a linearization for P and $\text{rev}L$ is a linearization for $\text{rev}P$.

THEOREM 3.3 (right eigenvector recovery from \mathbb{L}_2). *Let $L \in \mathbb{L}_2(P)$ be a linearization, with vector \tilde{v} (necessarily nonzero) in (3.4). If z is a right eigenvector of L with eigenvalue (α, β) then*

$$(3.6) \quad x = (\tilde{v}^T \otimes I_n)z$$

is a right eigenvector of P with eigenvalue (α, β) . Moreover, any right eigenvector of P corresponding to (α, β) can be recovered from one of L from the formula (3.6).

Proof. The proof is entirely analogous to that of Theorem 3.2. \square

The broader significance of Theorems 3.2 and 3.3 combined with Theorem 3.1 is that both left and right eigenvectors of pencils in \mathbb{L}_1 and \mathbb{L}_2 yield corresponding eigenvectors of P via simple formulae.

We will not write down backward error bounds for \mathbb{L}_1 and \mathbb{L}_2 , but will do so for their intersection in section 3.4.

3.2. Companion linearizations. It is easy to see that $C_2(P) = C_1(P^T)^T$, where P^T denotes the polynomial obtained by transposing each coefficient matrix A_i . This property implies that any backward error results for C_1 have a counterpart for C_2 , and so it suffices to concentrate on the first companion form.

Is the factorization (2.11) possible for the first companion linearization? For C_1 in (1.3) with $m = 3$ it is straightforward to verify that $E(\alpha, \beta)C_1(\alpha, \beta) = I_3 \otimes P(\alpha, \beta)$ with

$$E(\alpha, \beta) = \begin{bmatrix} \alpha^2 I_n & -(\beta^2 A_0 + \alpha\beta A_1) & -\alpha\beta A_0 \\ \alpha\beta I_n & \alpha\beta A_2 + \alpha^2 A_3 & -\beta^2 A_0 \\ \beta^2 I_n & \beta^2 A_2 + \alpha\beta A_3 & \beta^2 A_1 + \alpha\beta A_2 + \alpha^2 A_3 \end{bmatrix}$$

(indeed the first block row of this equation was mentioned in section 2.2), so that we have three choices for $G(\alpha, \beta)$, namely, $G_k(\alpha, \beta) := (e_k^T \otimes I_n)E(\alpha, \beta)$, $k = 1:3$. This result generalizes to arbitrary degrees m .

LEMMA 3.4. *For the first companion form $C_1(\alpha, \beta) = \alpha X_1 + \beta Y_1$, for any m , there exists a block $m \times m$ matrix $E(\alpha, \beta) \in \mathbb{C}^{mn \times mn}$ such that*

$$(3.7) \quad E(\alpha, \beta)C_1(\alpha, \beta) = I_m \otimes P(\alpha, \beta),$$

where the blocks are given by

$$[E(\alpha, \beta)]_{i1} = \alpha^{m-i} \beta^{i-1} I_n, \quad [E(\alpha, \beta)]_{ij} = \sum_{k=0}^{m-1} s_k \alpha^k \beta^{m-k-1} A_{\ell_k} \text{ for } j > 1,$$

where $s_k \in \{-1, 0, 1\}$ and the indices ℓ_k are distinct (our notation suppresses the dependence of s_k and ℓ_k on i and j). The condition (2.11) is satisfied for

$$(3.8) \quad G_k(\alpha, \beta) = (e_k^T \otimes I_n)E(\alpha, \beta), \quad g = e_k, \quad k = 1:m.$$

Proof. The proof consists of a direct verification that $E(\alpha, \beta)$ defined by

$$[E(\alpha, \beta)]_{ij} = \begin{cases} \alpha^{m-i} \beta^{i-1} I_n, & 1 \leq i \leq m, j = 1, \\ -(\alpha/\beta)^{j-i} \sum_{k=0}^{m-j} \alpha^{k-1} \beta^{m-k} A_k, & 1 \leq i < j, 1 < j \leq m, \\ (\alpha/\beta)^{j-i} \sum_{k=m-j+1}^m \alpha^{k-1} \beta^{m-k} A_k, & 1 < j \leq i \leq m, \end{cases}$$

satisfies (3.7). \square

The next lemma will be useful when taking norms of block matrices.

LEMMA 3.5. *For any block $\ell \times m$ matrix B we have $\|B\|_2 \leq \sqrt{\ell m} \max_{i,j} \|B_{ij}\|_2$.*

Proof. Partitioning x conformably with B , we have

$$\begin{aligned} \|Bx\|_2^2 &= \sum_{i=1}^{\ell} \left\| \sum_{j=1}^m B_{ij}x_j \right\|_2^2 \leq \max_{i,j} \|B_{ij}\|_2^2 \sum_{i=1}^{\ell} \left(\sum_{j=1}^m \|x_j\|_2 \right)^2 \\ &\leq \max_{i,j} \|B_{ij}\|_2^2 \sum_{i=1}^{\ell} m \left(\sum_{j=1}^m \|x_j\|_2^2 \right) = \ell m \max_{i,j} \|B_{ij}\|_2^2 \|x\|_2^2. \end{aligned}$$

The result follows. \square

To investigate the size of the upper bound in (2.14) for $L(\alpha, \beta) = C_1(\alpha, \beta) = \alpha X_1 + \beta Y_1$ we need to bound $\|X_1\|_2$, $\|Y_1\|_2$, and the norm of the k th block row $G_k(\alpha, \beta)$ of $E(\alpha, \beta)$. We find that

$$(3.9) \quad \|X_1\|_2 = \max(\|A_m\|_2, 1), \quad \|Y_1\|_2 \leq m \max\left(1, \max_{i=0:m-1} \|A_i\|_2\right),$$

where we used Lemma 3.5 for Y_1 . From Lemma 3.4 we have, for $j > 1$,

$$\|E(\alpha, \beta)_{ij}\|_2 \leq \max_{\ell} \|A_{\ell}\|_2 \sum_{k=0}^{m-1} |\alpha|^k |\beta|^{m-k-1} = \|A_{\alpha, \beta}\|_1 \max_{\ell} \|A_{\ell}\|_2,$$

so that on using Lemma 3.5,

$$(3.10) \quad \|G_k(\alpha, \beta)\|_2 \leq \sqrt{m} \|A_{\alpha, \beta}\|_1 \max(1, \max_i \|A_i\|_2),$$

this upper bound being independent of k . We can now bound the ratio $\eta_P(z_k, \alpha, \beta) / \eta_L(z, \alpha, \beta)$ in terms of the approximate right eigenpair (z, α, β) and the coefficient matrices defining P .

THEOREM 3.6. *Let z be an approximate right eigenvector of C_1 corresponding to the approximate eigenvalue (α, β) . Then for $z_k = z((k-1)n + 1:kn)$, $k = 1:m$, we have*

$$(3.11) \quad \begin{aligned} \frac{1}{m^{1/2}} \leq \frac{\eta_P(z_k, \alpha, \beta)}{\eta_{C_1}(z, \alpha, \beta)} &\leq m^{3/2} \frac{(|\alpha| + |\beta|) \|A_{\alpha, \beta}\|_1 \max(1, \max_i \|A_i\|_2)^2 \|z\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2 \|z_k\|_2} \\ &\leq m^{5/2} \frac{\max(1, \max_i \|A_i\|_2)^2 \|z\|_2}{\min(\|A_0\|_2, \|A_m\|_2) \|z_k\|_2}. \end{aligned}$$

Proof. The first upper bound is obtained by combining (2.14) with (3.9) and (3.10). For the second upper bound it suffices to note that

$$\begin{aligned} \frac{(|\alpha| + |\beta|) \|A_{\alpha, \beta}\|_1}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2} &\leq \frac{(|\alpha| + |\beta|)(|\alpha|^{m-1} + |\alpha|^{m-2}|\beta| + \dots + |\beta|^{m-1})}{\min(\|A_0\|_2, \|A_m\|_2)(|\alpha|^m + |\beta|^m)} \\ &\leq \frac{m}{\min(\|A_0\|_2, \|A_m\|_2)} \end{aligned}$$

by [7, Lem. A.1, (A.1)]. To prove the lower bound, let $\{\Delta A_i\}$ be an optimal set of perturbations in the definition of η_P . These trivially yield feasible perturbations

$\Delta X_1 = \text{diag}(\Delta A_m, 0, \dots, 0)$ of X_1 and ΔY_1 of Y_1 , with ΔY_1 being zero except for the first block row $[\Delta A_{m-1}, \dots, \Delta A_0]$. $\|\Delta X_1\|_2 \leq \eta_P \|X_1\|_2$ is immediate. Using Lemma 3.5,

$$\|\Delta Y_1\|_2 \leq m^{1/2} \max_{i=0:m-1} \|\Delta A_i\|_2 \leq m^{1/2} \eta_P \max_{i=0:m-1} \|A_i\|_2 \leq m^{1/2} \eta_P \|Y_1\|_2. \quad \square$$

The theorem reveals two main sufficient conditions for η_P to be not much larger than η_{C_1} . The first is that $\|z\|_2/\|z_k\|_2$ is not much larger than 1. In the context of floating point arithmetic this requirement is to be expected, because if $\|z\|_2 \gg \|z_k\|_2$ then z_k is likely to have suffered damaging subtractive cancellation in its formation. The second condition is that $\min(\|A_0\|_2, \|A_m\|_2) \approx \max_i \|A_i\|_2 \approx 1$, which is certainly true if $\|A_i\|_2 \approx 1$ for all i . Since $C_1 \in \mathbb{L}_1(P)$, Theorem 3.1 shows that the exact eigenvector is of the form $z = A_{\alpha,\beta} \otimes x$; since the largest element of $A_{\alpha,\beta}$ is the first or the last we can achieve $\|z\|_2/\|z_k\|_2 \in [1, \sqrt{m}]$ by taking $k = 1$ if $|\alpha| \geq |\beta|$ or $k = m$ if $|\alpha| \leq |\beta|$. The importance for achieving a good backward error of recovering x from the largest block component of z has already been noted and shown empirically for the QEP by Tisseur [22, sec. 3.2]; our analysis provides theoretical confirmation for all degrees m .

We now turn to the backward error for a left eigenpair. Since $C_1 \in \mathbb{L}_1(P)$ with $v = e_1$ we have $L(\alpha, \beta)(A_{\alpha,\beta} \otimes I_n) = e_1 \otimes P(\alpha, \beta)$, so that (2.15) is satisfied with $H(\alpha, \beta) = A_{\alpha,\beta} \otimes I_n$ and $h = e_1$. The ensuing eigenvector recovery property is, from (2.17) or (3.5), $y = w(1:n)$. Before obtaining a backward error bound we give a more complete description of the relation between y and w , which will aid in the interpretation of the bound. The following result extends [7, Lem. 7.2], which is stated for simple, finite, nonzero eigenvalues, to an arbitrary eigenvalue expressed in (α, β) -form.

LEMMA 3.7 (left eigenvector recovery for C_1). *The vector $y \in \mathbb{C}^n$ is a left eigenvector of P corresponding to the eigenvalue (α, β) if and only if*

$$(3.12) \quad w = \begin{cases} \begin{bmatrix} [\alpha^{m-1} I_n]^* \\ -[\alpha^{m-2} \beta A_{m-2} + \dots + \alpha \beta^{m-2} A_1 + \beta^{m-1} A_0]^* \\ -[\alpha^{m-2} \beta A_{m-3} + \dots + \alpha^2 \beta^{m-3} A_1 + \alpha \beta^{m-2} A_0]^* \\ \vdots \\ -[\alpha^{m-2} \beta A_0]^* \end{bmatrix} y, & \alpha \neq 0, \\ \begin{bmatrix} [\beta^{m-1} I_n]^* \\ [\alpha \beta^{m-2} A_m + \beta^{m-1} A_{m-1}]^* \\ \vdots \\ [\alpha^{m-1} A_m + \dots + \alpha \beta^{m-2} A_2 + \beta^{m-1} A_1]^* \end{bmatrix} y, & \beta \neq 0, \end{cases}$$

is a left eigenvector of C_1 corresponding to (α, β) . Every left eigenvector of C_1 with eigenvalue (α, β) is of the form (3.12) for some left eigenvector y of P . For a finite eigenvalue, an alternative representation of w is

$$w^* = y^* [I_n, \quad B_{m-2}, \quad \dots \quad B_1, \quad B_0],$$

where $(P(t) - P(\lambda))/(t - \lambda) = \sum_{i=0}^{m-1} B_i t^i$ and $B_i = B_i(\lambda)$.

Proof. Note first that the two different formulae in (3.12) (either of which can be obtained from the other by multiplying through by the conjugate of $(\alpha/\beta)^{m-1}$ or its reciprocal and using $y^* P(\alpha, \beta) = 0$) are needed because when $\alpha = 0$ (and hence

$y^* A_0 = 0$), the first expression is zero, while when $\beta = 0$ (and hence $y^* A_m = 0$), the second expression is zero.

For the first part it suffices to note that for w as defined by (3.12) we have

$$w^* C_1(\alpha, \beta) = \begin{cases} y^* P(\alpha, \beta)(e_1^T \otimes I_n), & \alpha \neq 0, \\ y^* P(\alpha, \beta)(e_m^T \otimes I_n), & \alpha = 0. \end{cases}$$

For the next part, since C_1 is a strong linearization [4] and P is regular, any eigenvalue (α, β) of C_1 of geometric multiplicity k is also an eigenvalue of P of geometric multiplicity k . Any k linearly independent eigenvectors y of P for (α, β) clearly yield via (3.12) k linearly independent eigenvectors of L . Hence any eigenvector of L for (α, β) has the form (3.12).

The last part generalizes the analogous formula for scalar companion matrices given by Stewart [21, sec. 2]; we omit the proof. \square

Lemma 3.7 shows that even when the eigenvalue is multiple all the left eigenvectors of P can be obtained from the first n components of the left eigenvectors of C_1 .

We can now obtain the desired backward error bounds.

THEOREM 3.8. *Let w be an approximate left eigenvector of C_1 corresponding to the approximate eigenvalue (α, β) . Then for $w_1 = w(1:n)$ we have*

$$\begin{aligned} \frac{1}{m^{1/2}} &\leq \frac{\eta_P(w_1^*, \alpha, \beta)}{\eta_{C_1}(w^*, \alpha, \beta)} \leq m \frac{(|\alpha| + |\beta|) \|A_{\alpha, \beta}\|_2 \max(1, \max_i \|A_i\|_2) \|w\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2} \frac{\|w\|_2}{\|w_1\|_2} \\ (3.13) \qquad &\leq m^{3/2} \frac{\max(1, \max_i \|A_i\|_2) \|w\|_2}{\min(\|A_0\|_2, \|A_m\|_2) \|w_1\|_2}. \end{aligned}$$

Proof. The first upper bound follows directly from (2.18), (3.9), and $\|H_1(\alpha, \beta)\|_2 = \|A_{\alpha, \beta} \otimes I_n\|_2 = \|A_{\alpha, \beta}\|_2$. For the second upper bound it suffices to note that

$$\begin{aligned} \frac{(|\alpha| + |\beta|) \|A_{\alpha, \beta}\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2} &\leq \frac{(|\alpha| + |\beta|)(|\alpha|^{2(m-1)} + |\alpha|^{2(m-2)} |\beta|^2 + \dots + |\beta|^{2(m-1)})^{1/2}}{\min(\|A_0\|_2, \|A_m\|_2) (|\alpha|^m + |\beta|^m)} \\ (3.14) \qquad &\leq \frac{m^{1/2}}{\min(\|A_0\|_2, \|A_m\|_2)} \end{aligned}$$

by [7, Lem. A.1, (A.3)]. The proof of the lower bound is exactly the same as in Theorem 3.6. \square

Notice that compared with the bounds in Theorem 3.6 for right eigenpairs, the factor $\max_i \|A_i\|_2$ is not squared. However, k is no longer a free parameter and so the ratio $\|w\|_2/\|w_1\|_2$ is fixed. Theorem 3.8 shows that $\eta_P(w_1^*, \alpha, \beta) \approx \eta_{C_1}(w^*, \alpha, \beta)$ is guaranteed provided that $\min(\|A_0\|_2, \|A_m\|_2) \approx \max_i \|A_i\|_2 \approx 1$ and $\|w\|_2/\|w_1\|_2$ is not much larger than 1. If $\|A_i\|_2 \lesssim 1$ for all i then for an exact left eigenvector w the ratio $\|w\|_2/\|w_1\|_2$ is bounded by about $(m^3/3)^{1/2}$; this can be seen from the first equation in (3.12) if $|\alpha| \geq |\beta|$ and the second if $|\alpha| \leq |\beta|$.

A comparison with earlier work is instructive. Tisseur [22, Thm. 7] and Van Dooren and Dewilde [24, sec. 7] both show that solving a PEP by applying a backward stable solver to the first companion pencil is backward stable for the PEP, under certain conditions on the A_i . Van Dooren and Dewilde measure the perturbation ΔP to P by $\|[\Delta A_m, \dots, \Delta A_0]\|_F / \|[A_m, \dots, A_0]\|_F$ and show that $\|[A_m, \dots, A_0]\|_F = 1$ implies stability. Tisseur uses the more stringent measure $\max_i \|\Delta A_i\|_2 / \|A_i\|_2$, as in (2.1), and proves that $\|A_i\|_2 \equiv 1$ implies stability. These analyses are carried out without reference to specific eigenpairs or eigenvector recovery formulae and so they provide much less precise information than the bounds in Theorems 3.6 and 3.8.

3.3. Scaled companion linearizations. When the coefficient matrices of P have norms that differ widely, the companion matrices $C_i(\lambda)$, $i = 1, 2$, are badly scaled and the bounds of Theorems 3.6 and 3.8 signal that $\eta_P \gg \eta_{C_1}$ is possible. In this section we study the effect on the backward error of scaling the identity blocks of C_i .

Let $D = \text{diag}(d) \otimes I_n$, where $d \in \mathbb{R}^m$ with $d_1 = 1$ and $d_i > 0$, $i = 2:m$. It is easily checked that $DC_1(\lambda) \in \mathbb{L}_1(P)$ with $v = e_1$, that $C_2(\lambda)D \in \mathbb{L}_2(P)$ with $\tilde{v} = e_1$, and that both scaled companion pencils are always linearizations. Since $C_2(P)D = (DC_1(P^T))^T$ we can concentrate on DC_1 . The condition (2.11) becomes $G_k(\alpha, \beta)D^{-1} \cdot DC_1(\alpha, \beta) = e_k^T \otimes P(\alpha, \beta)$, where G_k is defined in (3.8), and we find that

$$(3.15) \quad \|DX_1\|_2 = \max\left(\max_{i>1} d_i, \|A_m\|_2\right),$$

$$(3.16) \quad \|DY_1\|_2 \leq m \max\left(\max_{i>1} d_i, \max_{i=0:m-1} \|A_i\|_2\right),$$

$$(3.17) \quad \|G_k(\alpha, \beta)D^{-1}\|_2 \leq \sqrt{m} \|A_{\alpha, \beta}\|_1 \max\left(1, \frac{\max_i \|A_i\|_2}{\min_{i>1} d_i}\right).$$

In particular, if we choose $d_i = \max_{\ell} \|A_{\ell}\|_2$, $i = 2:m$, then

$$\begin{aligned} \|DX_1\|_2 &= \max_i \|A_i\|_2, & \|DY_1\|_2 &\leq m \max_i \|A_i\|_2, \\ \|G_k(\alpha, \beta)D^{-1}\|_2 &\leq \sqrt{m} \|A_{\alpha, \beta}\|_1. \end{aligned}$$

As we now show, this scaling yields bounds for η_P/η_{DC_1} better than those for η_P/η_{C_1} . We introduce the quantity

$$(3.18) \quad \rho = \frac{\max_i \|A_i\|_2}{\min(\|A_0\|_2, \|A_m\|_2)},$$

which measures the scaling of the problem.

THEOREM 3.9. *Let $D_s = \text{diag}(1, s, \dots, s) \otimes I_n \in \mathbb{R}^{mn \times mn}$ with $s = \max_i \|A_i\|_2$. Let z and w be approximate right and left eigenvectors of $D_s C_1$ corresponding to the approximate eigenvalue (α, β) . Then for $z_k = z((k-1)n + 1:kn)$, $k = 1:m$, we have*

$$\frac{1}{m^{1/2}} \leq \frac{\eta_P(z_k, \alpha, \beta)}{\eta_{D_s C_1}(z, \alpha, \beta)} \leq m^{3/2} \frac{(|\alpha| + |\beta|) \|A_{\alpha, \beta}\|_1 \max_i \|A_i\|_2 \|z\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2 \|z_k\|_2} \leq m^{5/2} \rho \frac{\|z\|_2}{\|z_k\|_2},$$

and for $w_1 = w(1:n)$,

$$\frac{1}{m^{1/2}} \leq \frac{\eta_P(w_1^*, \alpha, \beta)}{\eta_{D_s C_1}(w^*, \alpha, \beta)} \leq m \frac{(|\alpha| + |\beta|) \|A_{\alpha, \beta}\|_2 \max_i \|A_i\|_2 \|w\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2 \|w_1\|_2} \leq m^{3/2} \rho \frac{\|w\|_2}{\|w_1\|_2}.$$

Proof. The proof is analogous to the proofs of Theorems 3.6 and 3.8, making use of (3.15)–(3.17). \square

The bounds of Theorem 3.9 for the scaled companion pencil improve upon those for the unscaled pencil in several ways.

1. For the right eigenvector, the term $\max(1, \max_i \|A_i\|_2^2) / \min(\|A_0\|_2, \|A_m\|_2)$ in (3.11) is replaced by ρ , which is much smaller if $\max_i \|A_i\|_2 \gg 1$ or $\max_i \|A_i\|_2 \ll 1$.

2. For the left eigenvector, the term $\max(1, \max_i \|A_i\|_2) / \min(\|A_0\|_2, \|A_m\|_2)$ in (3.13) is replaced by ρ , which is much smaller if $\max_i \|A_i\|_2 \ll 1$.

3. For the scaled companion pencil, $\|w\|_2/\|w_1\|_2$ is guaranteed to be $O(m^{3/2})$ for the exact eigenvector, as can be seen from the appropriate choice of formula in (3.12), bearing in mind that scaling changes w in (3.12) to $D_s^{-1}w$. To draw the same conclusion for the unscaled pencil we require $\max_i \|A_i\|_2 \lesssim 1$.

Our bounds suggest that scaling the identity blocks of C_1 can significantly improve the backward error of the recovered eigenvectors of P . We can, of course, employ more sophisticated two-sided scalings, including balancing [16], [25]. However, these scalings produce a new pencil not belonging to \mathbb{L}_1 , so our backward error bounds are not applicable to them.

3.4. $\mathbb{DL}(P)$ linearizations. From section 2.2 and the definition of \mathbb{L}_1 in (3.3), it is clear that for pencils $L \in \mathbb{L}_1$ our analysis provides upper bounds for the backward error η_P associated with approximate *left* eigenvectors of P recovered from approximate *left* eigenvectors of L . The same is true for $L \in \mathbb{L}_2$ and approximate *right* eigenvectors. We now concentrate on the intersection

$$(3.19) \quad \mathbb{DL}(P) = \mathbb{L}_1(P) \cap \mathbb{L}_2(P),$$

since for pencils in $\mathbb{DL}(P)$ we can obtain backward error bounds for both left and right eigenvectors. $\mathbb{DL}(P)$ is a much smaller space than $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$, being just m -dimensional. Indeed, it is shown in [19, Thm. 5.3] and [6, Thm. 3.4] that $L \in \mathbb{DL}(P)$ if and only if L satisfies the conditions in (3.3) and (3.4) with $\tilde{v} = v$. The general form of $\mathbb{DL}(P)$ for the quadratic $P(\lambda) = \lambda^2 A_2 + \lambda A_1 + A_0$ is given by

$$\mathbb{DL}(P) = \left\{ L(\lambda) = \lambda \begin{bmatrix} v_1 A_2 & v_2 A_2 \\ v_2 A_2 & v_2 A_1 - v_1 A_0 \end{bmatrix} + \begin{bmatrix} v_1 A_1 - v_2 A_2 & v_1 A_0 \\ v_1 A_0 & v_2 A_0 \end{bmatrix} : v \in \mathbb{C}^2 \right\},$$

which illustrates the fact that the companion pencils are not contained in $\mathbb{DL}(P)$ for any m . Just as for \mathbb{L}_1 and \mathbb{L}_2 , almost all pencils in $\mathbb{DL}(P)$ are linearizations [19, Thm. 6.8]. In fact, there is a beautiful characterization of the subset of pencils $L \in \mathbb{DL}(P)$ that are linearizations [19, Thm. 6.7]: they are those for which no eigenvalue of P is a root of the polynomial $p(\lambda; v) := v^T A = \sum_{i=1}^m v_i \lambda^{m-i}$, where when $v_1 = 0$ we define ∞ to be a root of $p(\lambda; v)$. Throughout this section we assume that the pencils $L \in \mathbb{DL}(P)$ under consideration are linearizations.

For pencils $L \in \mathbb{DL}(P)$, we have, by definition,

$$L(\alpha, \beta)(A_{\alpha, \beta} \otimes I_n) = v \otimes P(\alpha, \beta), \quad (A_{\alpha, \beta}^T \otimes I_n)L(\alpha, \beta) = v^T \otimes P(\alpha, \beta),$$

so that (2.11) and (2.15) hold with $G(\alpha, \beta) = A_{\alpha, \beta}^T \otimes I_n$ and $H(\alpha, \beta) = A_{\alpha, \beta} \otimes I_n$. Moreover, $\|G(\alpha, \beta)\|_2 = \|H(\alpha, \beta)\|_2 = \|A_{\alpha, \beta}\|_2$. From [7, Lem. 4.1] we know that $L(\alpha, \beta) = \alpha X + \beta Y$ satisfies

$$(3.20) \quad \max(\|X\|_2, \|Y\|_2) \leq m r^{1/2} \max_i \|A_i\|_2,$$

where r is the number of nonzeros in v and we assume $\|v\|_2 = 1$ without loss of generality. We now have the ingredients to obtain a backward error bound. Recall that ρ is defined in (3.18).

THEOREM 3.10. *Let $L \in \mathbb{DL}(P)$ with vector v in (3.3) be a linearization, where v has unit 2-norm and r nonzeros. Let z be an approximate right eigenvector of L corresponding to the approximate eigenvalue (α, β) . Then for $x = \sum_{i=1}^m v_i z_i$, where $z_i = z((i-1)n+1:in)$, we have*

$$\frac{\eta_P(x, \alpha, \beta)}{\eta_L(z, \alpha, \beta)} \leq m r^{1/2} \frac{(|\alpha| + |\beta|) \|A_{\alpha, \beta}\|_2 \max_i \|A_i\|_2 \|z\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2} \frac{\|z\|_2}{\|x\|_2} \leq m^{3/2} r^{1/2} \rho \frac{\|z\|_2}{\|x\|_2}.$$

Proof. Combine (2.14), (3.20), and (3.14). \square

Note that the exact z has the form $A_{\alpha,\beta} \otimes \xi$ so that $\|z\|_2 = \|A_{\alpha,\beta}\|_2 \|\xi\|_2$, and $\|x\|_2 = \|\sum_{i=1}^m v_i z_i\|_2 = |A_{\alpha,\beta}^T v| \|\xi\|_2$. Hence $\|z\|_2/\|x\|_2 = \|A_{\alpha,\beta}\|_2/|\mathbf{p}(\alpha, \beta; v)|$, where $\mathbf{p}(\alpha, \beta; v) = A_{\alpha,\beta}^T v = \sum_{i=1}^m v_i \alpha^{m-i} \beta^{i-1}$. Thus $\min\{\|z\|_2/\|x\|_2 : \|v\|_2 = 1\} = 1$, with equality attained for $v_* = \overline{A_{\alpha,\beta}}/\|A_{\alpha,\beta}\|_2$. This choice of v minimizes the second upper bound of Theorem 3.10. However, simply choosing $v = e_k$ where $\|z_k\|_2 = \max_{i=1:m} \|z_i\|_2$ ensures that $\|z\|_2/\|x\|_2 \leq \sqrt{m}$, which is perfectly adequate.

Intuitively, we might expect that $\eta_P(x, \alpha, \beta) \geq \eta_L(z, \alpha, \beta)$, at least to within some constant factor, but this is not necessarily the case. Consider, for example, the pencil

$$L(\lambda) = \lambda \begin{bmatrix} A & A \\ A & B - C \end{bmatrix} + \begin{bmatrix} B - A & C \\ C & C \end{bmatrix} \in \mathbb{DL}(\lambda^2 A + \lambda B + C),$$

which corresponds to $v = [1 \ 1]^T$ in (3.3). Suppose $A = B = I$ and $C = \epsilon I$ with $0 < \epsilon \ll 1$, and let $\Delta A = \delta I$ and $\Delta B = \Delta C = 0$. These perturbations have relative size $\max(\|\Delta A\|_2/\|A\|_2, \|\Delta B\|_2/\|B\|_2, \|\Delta C\|_2/\|C\|_2) = \delta$, but for the pencil $\max(\|\Delta X\|_2/\|X\|_2, \|\Delta Y\|_2/\|Y\|_2) \approx \max(\delta, \delta/\epsilon) = \delta/\epsilon$. Hence a small perturbation to P does not necessarily correspond to a small perturbation to L and η_P/η_L cannot therefore be bounded below by a positive constant. This phenomenon is not present for the pencils corresponding to $v = e_k$, which form the standard basis for $\mathbb{DL}(P)$ [6], because for these pencils each block of X and Y is plus or minus a single block A_i . We now specialize Theorem 3.10 to these pencils.

COROLLARY 3.11. *Let $L_k \in \mathbb{DL}(P)$ corresponding to $v = e_k$ in (3.3) be a linearization. Let z be an approximate right eigenvector of L_k corresponding to the approximate eigenvalue (α, β) . Then for $z_k = z((k-1)n+1:kn)$, we have*

$$\frac{1}{m} \leq \frac{\eta_P(z_k, \alpha, \beta)}{\eta_{L_k}(z, \alpha, \beta)} \leq m \frac{(|\alpha| + |\beta|) \|A_{\alpha,\beta}\|_2 \max_i \|A_i\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2} \frac{\|z\|_2}{\|z_k\|_2} \leq m^{3/2} \rho \frac{\|z\|_2}{\|z_k\|_2}.$$

Proof. The upper bound follows from Theorem 3.10. The lower bound is proved in a similar way to the lower bound of Theorem 3.6. \square

Analogues of Theorem 3.10 and Corollary 3.11 hold for approximate left eigenvectors w of L_k : z is simply replaced by w and x by $y = \sum_{i=1}^m \bar{v}_i w_i$.

With the notation in Corollary 3.11, the exact eigenvector z satisfies $z = A_{\alpha,\beta} \otimes x$, and it is easy to see that $\|z\|_2/\|z_k\|_2 \approx 1$ for $k = 1$ if $|\alpha| \geq |\beta|$ and for $k = m$ if $|\alpha| \leq |\beta|$. Assuming the approximate eigenvector z shares the latter property, the pencils in $\mathbb{DL}(P)$ with $v = e_1$ and $v = e_m$ yield backward errors $\eta_P \approx \eta_L$ for eigenpairs with eigenvectors of modulus greater than or less than 1, respectively, provided that the measure ρ of the scaling of the problem is of order 1. Two points are worth noting.

1. Although an eigenvector of P can be recovered from any of the blocks $z_i = z((i-1)n+1:in)$ of an eigenvector z of L_k (see Theorem 3.1), our backward error bounds in Corollary 3.11 require $i = k$.

2. The pencils L_1 and L_m are indeed linearizations if A_0 and A_m , respectively, are nonsingular, as can be seen from the characterization mentioned at the start of this subsection.

3.5. Comparison with conditioning results. Backward error and conditioning are complementary concepts. Ideally, we would like the linearization L that we use to be as well conditioned as the original polynomial P and for it to lead, after recovering an approximate eigenpair of P from one of L , to a backward error η_P

of the same order of magnitude as η_L . Therefore to show that one linearization is preferable to another we need to show that it enjoys a better bound for η_P/η_L as well as a better condition number bound. Remarkably, our backward error results are entirely harmonious with the results of Higham, Mackey, and Tisseur [7] concerning eigenvalue conditioning, as we now explain.

For the companion forms the analysis in [7, sec. 7] provides bounds on the ratio $\kappa_{C_1}(\lambda)/\kappa_P(\lambda)$ of appropriately defined condition numbers in the case of quadratics. That analysis is readily extended to (α, β) -form and general degrees, and it shows that, like the backward error ratio in Theorem 3.6, $\kappa_{C_1}(\alpha, \beta)/\kappa_P(\alpha, \beta)$ is bounded by a multiple of $\max(1, \max_i \|A_i\|_2^2)/\min(\|A_0\|_2, \|A_m\|_2)$. Thus if $\min(\|A_0\|_2, \|A_m\|_2) \approx \max_i \|A_i\|_2 \approx 1$ and if a relatively large block is used for right eigenvector recovery then C_1 is an optimal linearization from the points of view of both backward error and conditioning.

For the scaled companion forms we can show that $\kappa_{D_s C_1}(\alpha, \beta)/\kappa_P(\alpha, \beta)$ is bounded by a multiple of ρ , just as for the backward error ratios in Theorem 3.9. So if $\rho \approx 1$ and a relatively large block is used for eigenvector recovery then $D_s C_1$ is an optimal linearization.

For the $\mathbb{DL}(P)$ pencils with $v = e_1$ (if $|\lambda| \geq 1$) or $v = e_m$ (if $|\lambda| \leq 1$) it is once again the case that the factor ρ in the backward error bound (in Corollary 3.11) is also the key quantity in a bound on the ratio of condition numbers κ_{L_k}/κ_P . We can conclude that if $\rho \approx 1$ then L_1 and L_m are optimal with respect to both backward error and conditioning over all linearizations for $|\lambda|$ greater than 1 and less than 1, respectively, assuming $\|z_1\|_2 \approx \|z\|_2$ for L_1 and $\|z_m\|_2 \approx \|z\|_2$ for L_m (properties that hold for the exact eigenvectors).

4. Structured linearizations. We now briefly consider to what extent the results above extend to structured linearizations for structured polynomials. Our definition of backward error remains the same and so does not incorporate structure. The issue is that structure may change some key properties of a linearization and thereby may limit our freedom in choosing how to recover eigenvectors.

4.1. Symmetric and Hermitian structures. If P is symmetric, that is, $P(\lambda) = P(\lambda)^T$, then all the pencils in $\mathbb{DL}(P)$ are symmetric, and these comprise all the symmetric pencils in $\mathbb{L}_1(P)$ [6, Thm 5.2]. Hence Theorem 3.10 and Corollary 3.11 are both applicable with L symmetric. If P is Hermitian, that is, $P(\lambda) = P(\bar{\lambda})^*$, then it is precisely the pencils in $\mathbb{DL}(P)$ with a *real* vector v that are Hermitian [6, Thm 6.2]. Theorem 3.10 remains applicable for Hermitian L with the minor restriction that v is real. Thus symmetry and Hermitian structure impose no significant limitations on the applicability of our backward error bounds.

4.2. Alternating and palindromic structures. We now consider some other classes of structures for which we can identify structured linearizations. These structures are less familiar than symmetric or Hermitian structures but still important in a variety of applications [17, Chap. 7]. In what follows, the symbol \star is used as an abbreviation for transpose (T) in the real case and either transpose or conjugate transpose ($*$) in the complex case. The \star -adjoint of P is defined by

$$P^\star(\lambda) = \sum_{i=0}^m \lambda^i A_i^\star.$$

$P(\lambda)$ is said to be

★-even if $P^*(-\lambda) = P(\lambda)$, ★-odd if $P^*(-\lambda) = -P(\lambda)$,
 ★-palindromic if $\text{rev}P^*(\lambda) = P(\lambda)$, ★-antipalindromic if $\text{rev}P^*(\lambda) = -P(\lambda)$,

where rev is defined in (3.2). For example, the quadratic $Q(\lambda) = \lambda^2 M + \lambda G + K$ with M, K symmetric and G skew-symmetric, arising in gyroscopic systems, is T -even since $Q^T(-\lambda) = Q(\lambda)$. On the other hand, the quadratic $Q(\lambda) = \lambda^2 A + \lambda B + A^T$ with B complex symmetric, arising in the study of vibration of rail tracks under the excitation of high speed trains [10], [11], is T -palindromic since $\text{rev}Q^T(\lambda) = Q(\lambda)$.

Linearizations in $\mathbb{L}_1(P)$ that reflect the structure of these polynomials and therefore preserve symmetries in their spectra have recently been investigated by Mackey et al. [18]. It is shown in [18, Thms. 3.5, 3.6] that if $L(\lambda) \in \mathbb{L}_1(P)$ is ★-structured with vector v then $(M \otimes I_n)L(\lambda)$ is in $\mathbb{DL}(P)$ with vector Mv , where M is either a diagonal matrix of alternating signs, $M = \text{diag}((-1)^{m-1}, \dots, (-1)^0)$, in the case of even/odd structures, or the reverse identity matrix, $R = (\delta_{i,n+1-i})$, in the context of palindromic structures.

Since L itself is in general not in $\mathbb{DL}(P)$ we cannot apply Theorem 3.10. However, the proof of the theorem is readily adapted, and by exploiting the fact that $M \otimes I_n$ is unitary the same bound is obtained.

THEOREM 4.1. *Let $L \in \mathbb{L}_1(P)$ with vector v be a ★-structured linearization and assume that v has unit 2-norm and r nonzeros. Let z be an approximate right eigenvector of L corresponding to the approximate eigenvalue (α, β) . Then for $x = \sum_{i=1}^m v_i z_i$ we have*

$$\frac{\eta_P(x, \alpha, \beta)}{\eta_L(z, \alpha, \beta)} \leq m r^{1/2} \frac{(|\alpha| + |\beta|) \|A_{\alpha, \beta}\|_2 \max_i \|A_i\|_2 \|z\|_2}{\sum_{i=0}^m |\alpha|^i |\beta|^{m-i} \|A_i\|_2 \|x\|_2} \leq m^{3/2} r^{1/2} \rho \frac{\|z\|_2}{\|x\|_2}.$$

For approximate left eigenvectors an analogous bound holds with z replaced by w and x by $y = \sum_{i=1}^m (M\bar{v})_i w_i$.

Theorem 4.1 shows that $\eta_P \approx \eta_L$ as long as $\rho = O(1)$ and $\|z\|_2/\|x\|_2 \approx 1$. However, whereas for $\mathbb{DL}(P)$ v can be freely chosen, in particular to minimize $\|z\|_2/\|x\|_2$, now the choice of v is constrained by the requirement that L be ★-structured. For example, for T -palindromic polynomials P , $L \in \mathbb{L}_1(P)$ with vector v is T -palindromic if and only $Rv = v$ [18, Thm. 3.5]; in the case of a quadratic, $v = [1 \ 1]/\sqrt{2}$ is forced.

5. Quadratic polynomials. We now concentrate our attention on quadratic polynomials, $Q(\lambda) = \lambda^2 A + \lambda B + C$, for which we can give a more detailed analysis than in the general case, covering in particular a potentially very beneficial scaling of the polynomial. We write

$$(5.1) \quad a = \|A\|_2, \quad b = \|B\|_2, \quad c = \|C\|_2.$$

Note that $A_{\alpha, \beta} = [\alpha, \beta]^T$. We will recover eigenvectors of Q from the components $z_1 = z(1:n)$ and $z_2 = z(n+1:2n)$ (and similarly for w) of eigenvectors of a linearization.

The first companion form of Q is given by

$$C_1(\lambda) = \lambda \begin{bmatrix} A & 0 \\ 0 & I_n \end{bmatrix} + \begin{bmatrix} B & C \\ -I_n & 0 \end{bmatrix},$$

and $D_s C_1(\lambda) = \text{diag}(I_n, sI_n) C_1(\lambda)$ with $s = \max(a, b, c)$. We normalize so that $|\alpha|^2 + |\beta|^2 = 1$. Theorems 3.6 and 3.9 say that for right eigenpairs

$$(5.2) \quad \frac{\eta_Q(z_k, \alpha, \beta)}{\eta_{C_1}(z, \alpha, \beta)} \leq 2^{5/2} \frac{\max(1, a, b, c)^2}{|\alpha|^2 a + |\alpha| |\beta| b + |\beta|^2 c} \frac{\|z\|_2}{\|z_k\|_2}, \quad k = 1, 2,$$

$$(5.3) \quad \frac{\eta_Q(z_k, \alpha, \beta)}{\eta_{D_s C_1}(z, \alpha, \beta)} \leq 2^{5/2} \frac{\max(a, b, c)}{|\alpha|^2 a + |\alpha||\beta|b + |\beta|^2 c} \frac{\|z\|_2}{\|z_k\|_2}, \quad k = 1, 2.$$

Analogous bounds hold for left eigenvectors: they have factor $2^{3/2}$ and there is no square in the numerator for the analogue of (5.2). In interpreting these bounds and those below recall that, for the exact eigenvectors of any pencil in $\mathbb{L}_1(Q)$,

$$(5.4) \quad \frac{\|z\|_2}{\|z_1\|_2} \approx 1 \quad \text{for } |\alpha| \geq |\beta|, \quad \frac{\|z\|_2}{\|z_2\|_2} \approx 1 \quad \text{for } |\alpha| \leq |\beta|.$$

The $\mathbb{DL}(Q)$ pencils with $v = e_1$ and $v = e_2$ are given by

$$L_1(\lambda) = \lambda \begin{bmatrix} A & 0 \\ 0 & -C \end{bmatrix} + \begin{bmatrix} B & C \\ C & 0 \end{bmatrix}, \quad L_2(\lambda) = \lambda \begin{bmatrix} 0 & A \\ A & B \end{bmatrix} + \begin{bmatrix} -A & 0 \\ 0 & C \end{bmatrix}.$$

We know from Corollary 3.11 that

$$(5.5) \quad \frac{\eta_Q(z_1, \alpha, \beta)}{\eta_{L_1}(z, \alpha, \beta)} \leq 2^{3/2} \frac{\max(a, b, c)}{|\alpha|^2 a + |\alpha||\beta|b + |\beta|^2 c} \frac{\|z\|_2}{\|z_1\|_2}.$$

In view of (5.4) this bound is appropriate when $|\alpha| \geq |\beta|$. If $|\alpha| \leq |\beta|$ then we wish to take z_2 rather than z_1 as eigenvector of Q , but Theorem 3.10 does not provide a bound for L_1 and z_2 . We now derive such a bound, by explicitly constructing an appropriate G matrix. It is easy to check that $G_Q(\alpha, \beta) = [\beta I_n, \quad -(\alpha A + \beta B)C^{-1}]$ satisfies $G_Q(\alpha, \beta)L_1(\alpha, \beta) = e_2^T \otimes Q(\alpha, \beta)$ so that (2.11) holds, and by Lemma 3.5

$$\|G_Q(\alpha, \beta)\|_2 \leq \sqrt{2} \|A_{\alpha, \beta}\|_\infty \max(1, (a + b)\|C^{-1}\|_2).$$

Hence (2.14) yields

$$(5.6) \quad \frac{\eta_Q(z_2, \alpha, \beta)}{\eta_{L_1}(z, \alpha, \beta)} \leq 4 \frac{\max(a, b, c) \max(1, (a + b)\|C^{-1}\|_2)}{|\alpha|^2 a + |\alpha||\beta|b + |\beta|^2 c} \frac{\|z\|_2}{\|z_2\|_2}.$$

Similarly we have for L_2 , by an analogue of the G_Q analysis and by Corollary 3.11,

$$(5.7) \quad \frac{\eta_Q(z_1, \alpha, \beta)}{\eta_{L_2}(z, \alpha, \beta)} \leq 4 \frac{\max(a, b, c) \max(1, (b + c)\|A^{-1}\|_2)}{|\alpha|^2 a + |\alpha||\beta|b + |\beta|^2 c} \frac{\|z\|_2}{\|z_1\|_2},$$

$$(5.8) \quad \frac{\eta_Q(z_2, \alpha, \beta)}{\eta_{L_2}(z, \alpha, \beta)} \leq 2^{3/2} \frac{\max(a, b, c)}{|\alpha|^2 a + |\alpha||\beta|b + |\beta|^2 c} \frac{\|z\|_2}{\|z_2\|_2}.$$

Essentially the same bounds (5.5)–(5.8) hold for approximate left eigenvectors: z is simply replaced by w and z_i by w_i .

In Table 5.1 we summarize for unstructured quadratics the main conclusions from these bounds concerning conditions that guarantee $\eta_P \approx \eta_L$. Here, using ρ from (3.18),

$$(5.9) \quad \rho = \frac{\max(a, b, c)}{\min(a, c)} \geq \frac{\max(a, b, c)}{|\alpha|^2 a + |\alpha||\beta|b + |\beta|^2 c}.$$

In view of the bounds, it is natural to scale the problem to try to bring the 2-norms of A , B , and C close to 1. The scaling of Fan, Lin, and Van Dooren [2] has

TABLE 5.1
Sufficient conditions for $\eta_P \approx \eta_L$; ρ is defined in (5.9).

Linearization	Eigenvalue	Right eigenvector	Left eigenvector	Condition
Companion	$ \alpha \geq \beta $ $ \alpha \leq \beta $	z_1 z_2	w_1	$b \leq a \approx c \approx 1$
Scaled companion	$ \alpha \geq \beta $ $ \alpha \leq \beta $	z_1 z_2	w_1	$\rho \approx 1$
L_1	$ \alpha \geq \beta $ $ \alpha \leq \beta $	z_1 z_2	w_1 w_2	$\rho \approx 1$ $\rho \max(1, (a+b)\ C^{-1}\ _2) \approx 1$
L_2	$ \alpha \geq \beta $ $ \alpha \leq \beta $	z_1 z_2	w_1 w_2	$\rho \max(1, (b+c)\ A^{-1}\ _2) \approx 1$ $\rho \approx 1$

precisely this aim. It converts $Q(\lambda) = \lambda^2 A + \lambda B + C$ to $\tilde{Q}(\mu) = \mu^2 \tilde{A} + \mu \tilde{B} + \tilde{C}$, where

$$(5.10a) \quad \lambda = \gamma\mu, \quad Q(\lambda)\delta = \mu^2(\gamma^2\delta A) + \mu(\gamma\delta B) + \delta C \equiv \tilde{Q}(\mu),$$

$$(5.10b) \quad \gamma = \sqrt{c/a}, \quad \delta = 2/(c + b\gamma).$$

Letting

$$(5.11) \quad \tilde{a} = \|\tilde{A}\|_2, \quad \tilde{b} = \|\tilde{B}\|_2, \quad \tilde{c} = \|\tilde{C}\|_2,$$

$$(5.12) \quad \tau = \frac{b}{\sqrt{ac}},$$

we have

$$\tilde{a} = \tilde{c} = \frac{2}{1 + \tau}, \quad \tilde{b} = \frac{2\tau}{1 + \tau}, \quad \frac{\tilde{a}}{2} + \tilde{b} + \frac{\tilde{c}}{2} = 2,$$

so that $2/3 \leq \max(\tilde{a}, \tilde{b}, \tilde{c}) \leq 2$. It is straightforward to show that $\tilde{\rho} = \max(\tilde{a}, \tilde{b}, \tilde{c}) / \min(\tilde{a}, \tilde{c}) = \max(1, \tau) \leq \rho$. Note that $\eta_Q(x, \lambda) = \eta_{\tilde{Q}}(x, \mu)$, so this scaling has no effect on the backward error for the quadratic; its purpose is to improve the backward error for the linearization. For $\tilde{Q}(\mu)$, the bounds (5.2), (5.3), and (5.5)–(5.8) can be simplified.

THEOREM 5.1. *Let (z, w, α, β) be an approximate eigentriple of a linearization of the scaled quadratic \tilde{Q} in (5.10) with $|\alpha|^2 + |\beta|^2 = 1$. Define*

$$(5.13) \quad \omega = \omega(\alpha, \beta) := \frac{1 + \tau}{1 + |\alpha\beta|\tau},$$

with τ as in (5.12). We have

$$\frac{\eta_{\tilde{Q}}(z_i, \alpha, \beta)}{\eta_{C_1}(z, \alpha, \beta)} \leq 2^{7/2} \omega \frac{\|z\|_2}{\|z_i\|_2}, \quad i = 1, 2, \quad \frac{\eta_{\tilde{Q}}(w_1^*, \alpha, \beta)}{\eta_{C_1}(w^*, \alpha, \beta)} \leq 2^{3/2} \omega \frac{\|w\|_2}{\|w_1\|_2}.$$

The same bounds hold for $D_s C_1$ and the constant $2^{7/2}$ can be replaced by $2^{5/2}$. Furthermore,

$$\frac{\eta_{\tilde{Q}}(x, \alpha, \beta)}{\eta_{L_i}(z, \alpha, \beta)} \leq f_i(x) \omega \frac{\|z\|_2}{\|x\|_2}, \quad i = 1, 2,$$

with

$$f_1(x) = \begin{cases} 2^{3/2} & \text{if } x = z_1, \\ 8\|\tilde{C}^{-1}\|_2 & \text{if } x = z_2, \end{cases} \quad f_2(x) = \begin{cases} 8\|\tilde{A}^{-1}\|_2 & \text{if } x = z_1, \\ 2^{3/2} & \text{if } x = z_2, \end{cases}$$

where the nonsingularity of C and A is required for $f_1(z_2)$ and $f_2(z_1)$. Similar bounds hold for approximate left eigenvectors and L_i , $i = 1, 2$: z is replaced by w and z_i by w_i .

Proof. For the scaled norms in (5.11) we have

$$(5.14) \quad |\alpha|^2\tilde{a} + |\alpha||\beta|\tilde{b} + |\beta|^2\tilde{c} = \frac{2}{1+\tau} + |\alpha||\beta|\frac{2\tau}{1+\tau} = \frac{2(1+|\alpha||\beta|\tau)}{1+\tau} = \frac{2}{\omega(\alpha, \beta)}$$

and the upper bounds follow from (5.2), (5.3), and (5.5)–(5.8). \square

We can regard ω in (5.13) as a growth factor bound in the translation from backward error for L to backward error for \tilde{Q} . With the normalization $|\alpha|^2 + |\beta|^2 = 1$, which implies $|\alpha||\beta| \leq 1/2$, this factor satisfies the bounds

$$(5.15) \quad 1 \leq \frac{1+\tau}{1+\frac{1}{2}\tau} \leq \omega \leq \min \left\{ 1+\tau, \frac{1}{|\alpha\beta|} \right\} \leq 1+\tau.$$

Fan, Lin, and Van Dooren [2] identify $\max(1+\tau, 1+\tau^{-1})$ as a growth factor. Our bounds for ω , which unlike in [2] are for individual eigenpairs and apply to L_i as well as C_1 , are sharper in two respects. First, they show that τ satisfying $\tau \ll 1$ are harmless, since our upper bound for ω is $O(1)$. Second, even when $\tau \gg 1$ the penultimate bound in (5.15) will still be of order 1 if $|\alpha||\beta| = |\alpha|\sqrt{1-|\alpha|^2} = O(1)$, which is the case unless $|\lambda| = |\alpha|/|\beta| = |\alpha|/\sqrt{1-|\alpha|^2}$ is small or large.

The most striking consequence of the theorem is that if

$$(5.16) \quad \|B\|_2 \lesssim (\|A\|_2\|C\|_2)^{1/2},$$

so that $\tau = O(1)$ and hence $\omega = O(1)$, then the η_Q/η_L ratios are 1 for the relevant choice of z_i , provided \tilde{A}^{-1} and \tilde{C}^{-1} have norms of order 1 in the case of two of the bounds for L_1 and L_2 . In the terminology of quadratics arising from mechanical systems with damping, the condition (5.16) holds for systems that are not too heavily damped. A class of problems for which (5.16) is satisfied is the elliptic Q [9], [13]: those for which A is Hermitian positive definite, B and C are Hermitian, and $(x^*Bx)^2 < 4(x^*Ax)(x^*Cx)$ for all nonzero $x \in \mathbb{C}^n$.

Our conclusions about the benefits to the backward error of scaling Q apply equally well to the condition numbers. Indeed, using (5.14) the analysis in [7] can be improved to provide bounds for $\kappa_L/\kappa_{\tilde{Q}}$ expressed in terms of ω instead of ρ for $L = C_1$, L_1 , and L_2 . Therefore for these three choices of L both backward error (modulo the potential requirement that $\|\tilde{A}^{-1}\|, \|\tilde{C}^{-1}\| = O(1)$ for L_1 and L_2) and conditioning are essentially optimal for the scaled problem if $\omega = O(1)$.

6. Numerical experiments. We illustrate the theory on three symmetric QEPs. Our experiments were performed in MATLAB 7, for which the unit roundoff is $u = 2^{-53} \approx 1.1 \times 10^{-16}$. The eigenpairs of $L(\lambda)$ were computed by MATLAB's function `qz`. Table 6.1 reports the problem sizes, the coefficient matrix norms, and the values of ρ in (3.18) (or (5.9)) before and after scaling via (5.10). In our figures, the x -axis is the eigenvalue index and the eigenvalues are sorted in increasing order

TABLE 6.1
Problem statistics. Here, $|\lambda_{\min}| = \min_i |\lambda_i|$, $|\lambda_{\max}| = \max_i |\lambda_i|$.

Problem	Wave		Nuclear		Mass-spring	
	25		8		50	
n	Unscaled	Scaled	Unscaled	Scaled	Unscaled	Scaled
$ \lambda_{\min} $	1.0e0	4.0e-2	1.8e1	6.6e-2	1.6e-2	7.0e-3
$ \lambda_{\max} $	2.5e1	1.0e0	3.6e2	1.4e0	3.2e2	1.4e2
$\ A\ _2$	1.6e0	1.9e0	2.4e8	1.2e0	1.0e0	1.4e-2
$\ B\ _2$	3.2e0	1.5e-1	4.4e10	8.2e-1	3.2e2	2.0e0
$\ C\ _2$	9.8e2	1.9e0	1.7e13	1.2e0	5.0e0	1.4e-2
$\ A^{-1}\ _2$	6.4e-1	5.4e-1	1.8e-1	3.7e7	1.0e0	7.2e1
$\ C^{-1}\ _2$	6.4e-1	3.4e2	2.1e-4	2.9e9	1.0e0	3.6e2
ρ	6.2e2	1.0e0	7.1e4	1.0e0	3.2e2	1.4e2
	$\tau = 8.0e-2, \max \omega = 1.1e0$		$\tau = 7.0e-1, \max \omega = 1.6e0$		$\tau = 1.4e2, \max \omega = 7.2e1$	

of absolute value. Throughout this section “companion” refers to the first companion linearization, C_1 .

Our first problem comes from applying the Galerkin method to a PDE describing the wave motion of a vibrating string with clamped ends in a spatially inhomogeneous environment [3], [9]. The quadratic Q is elliptic. Table 6.2 displays the smallest and largest ratios $\eta_Q(x, \alpha, \beta)/\eta_L(z, \alpha, b)$ over all computed eigenvalues for several linearizations and for the two ways of recovering the right eigenvector: $x = z_1$ and $x = z_2$. These ratios are compared with the corresponding theoretical upper bounds (5.2), (5.3), and (5.5)–(5.8) (taking the same (α, β) as for the smallest/largest backward error ratio). The upper bounds for the scaled companion linearization are smaller than those for the companion linearization, as expected by the theory since $c \gg 1$, and they also are sharper. For $\mathbb{DL}(Q)$ linearizations, the theory suggests using L_1 with $x = z_1$, since all the eigenvalues of Q have modulus at least 1. This is reflected in Table 6.2, where the L_1, z_1 pairing produces smaller ratios and upper bounds than L_2, z_1 . For the scaled quadratic \tilde{Q} in (5.10), we computed the bounds of Theorem 5.1. Since this problem is elliptic, we know from Theorem 5.1 that for the scaled problem, whose eigenvalues lie between 0.04 and 1 in modulus, the scaled and unscaled companion linearizations and the $\mathbb{DL}(\tilde{Q})$ linearization L_2 will have backward errors similar to those for \tilde{Q} for every eigenvalue with the choice $x = z_2$. This is confirmed by the boldface entries in the last two columns of Table 6.2.

Our second problem is a simplified model of a nuclear power plant, as described in [12], [23]. The largest ratios $\eta_Q(x, \alpha, \beta)/\eta_L(z, \alpha, b)$ and corresponding upper bounds are displayed in Table 6.3. Similar conclusions to those for the wave problem can be drawn for this problem. Since $\rho = 7 \times 10^4$, it is not surprising that some very large ratios are obtained. This example also illustrates the advantage of scaling the companion matrix. This is even more striking in Figure 6.1, where the ratios for all the right and left eigenpairs are displayed. For the companion linearization, these ratios can be up to 10^{10} times as large as those for $D_s C_1$. Although the problem is not elliptic, $\|B\|_2 \leq \sqrt{\|A\|_2 \|C\|_2}$ holds, and so our theory says that scaling will make the scaled and unscaled companion linearizations and the $\mathbb{DL}(Q)$ linearization L_2 with $x = z_2$ (since the scaled eigenvalues have modulus at most 1) optimally stable. This prediction is confirmed by the boldface entries in Table 6.3. Notice that for the scaled quadratic \tilde{Q} , the bounds for L_1 with z_2 and L_2 with z_1 are very weak, due to the large values of $\|\tilde{A}^{-1}\|_2$ and $\|\tilde{C}^{-1}\|_2$ shown in Table 6.1.

Our third problem is a standard damped mass-spring system, as described in [23,

TABLE 6.2
Wave problem, $n = 25$.

		Unscaled, $\rho = 6e2$				Scaled, $\rho = 1$	
Linearization L	Ei'vec x	$\min \frac{\eta_Q}{\eta_L}$	Upper bound	$\max \frac{\eta_Q}{\eta_L}$	Upper bound	$\max \frac{\eta_{\tilde{Q}}}{\eta_L}$	Upper bound
Companion	z_1	2.0e1	1.7e6	3.7e2	1.5e6	2.9e1	3.0e2
	z_2	1.4e0	1.6e4	3.8e2	3.5e7	2.5e0	1.6e1
Scaled companion	z_1	9.8e-1	1.6e1	1.1e2	1.7e3	1.4e1	1.5e2
	z_2	6.9e-1	1.6e1	1.1e2	4.3e4	1.7e0	8.3e0
L_1	z_1	1.9e0	2.2e1	9.0e1	1.1e3	1.8e1	7.6e1
	z_2	1.7e0	2.4e1	8.3e2	6.4e4	8.9e1	2.8e3
L_2	z_1	2.2e1	9.8e5	1.7e2	2.0e4	1.4e1	1.6e2
	z_2	1.5e0	8.0e0	1.5e2	2.1e4	1.7e0	8.3e0

TABLE 6.3
Nuclear problem, $n = 8$.

		Unscaled, $\rho = 7e4$				Scaled, $\rho = 1$	
Linearization L	Ei'vec x	$\min \frac{\eta_Q}{\eta_L}$	Upper bound	$\max \frac{\eta_Q}{\eta_L}$	Upper bound	$\max \frac{\eta_{\tilde{Q}}}{\eta_L}$	Upper bound
Companion	z_1	6.1e5	3.2e18	2.6e11	2.8e16	4.2e0	2.8e2
	z_2	5.5e5	1.2e21	9.3e9	9.9e19	3.1e-1	1.8e1
Scaled companion	z_1	3.4e-1	4.1e4	2.3e1	5.1e4	1.0e1	1.4e2
	z_2	3.0e-1	4.1e6	2.1e1	5.9e6	2.2e-1	9.2e0
L_1	z_1	2.1e1	1.2e3	2.3e3	1.4e5	1.4e1	6.9e1
	z_2	2.0e1	1.4e11	5.7e5	3.2e14	8.3e2	2.4e10
L_2	z_1	4.4e3	6.0e17	4.6e4	1.6e17	4.4e1	3.7e9
	z_2	1.2e2	1.5e4	1.9e3	6.2e6	7.6e-1	9.1e0

sec. 3.9]. The matrix $A = I$, B is tridiagonal with super- and subdiagonal elements all -64 and diagonal $128, 192, 192, \dots, 192$, and C is tridiagonal with super- and subdiagonal elements all -1 and diagonal $2, 3, \dots, 3$. The eigenvalues are all negative, with 50 eigenvalues of large modulus ranging from -320 to -6.4 and 50 small modulus eigenvalues approximately -1.5×10^{-2} . For the approximate right eigenvector, we take $x = z_1$ if $|\lambda| \geq 1$ and $x = z_2$ otherwise, as suggested by the theory. The largest ratios $\eta_Q(x, \alpha, \beta)/\eta_L(z, \alpha, b)$ and corresponding upper bounds are displayed in Table 6.4. Notice that for this problem the upper bound on the ratio $\eta_Q(x, \alpha, \beta)/\eta_{C_1}(z, \alpha, \beta)$ is nearly attained, which suggests that the factor $\max(1, a, b, c)^2$ in the bound should indeed contain the square. The largest ratio for $L = L_1$ corresponds to a small eigenvalue with $x = z_2$ and, for $L = L_2$, the largest ratio corresponds to a large eigenvalue with $x = z_1$. Hence, the reported upper bounds contain the extra factors $(a + b)\|C^{-1}\|_2$ and $(b + c)\|A^{-1}\|_2$, respectively, which explains why the bounds are larger than those for the scaled companion linearization (on the scaled and unscaled problems), which are small multiples of ρ . The top plot in Figure 6.2 shows that for the scaled quadratic \tilde{Q} , small backward error ratios are obtained for $L = L_1$ and large eigenvalues, whereas the ratios are small with the choice $L = L_2$ for the small eigenvalues—all as the theory predicts. The bottom plot in Figure 6.2 confirms that the actual backward errors η_Q are what we would expect, given the ratios and the fact that the computed eigenpairs of L are obtained via the QZ algorithm and so necessarily have a backward error of order u .

Finally, we mention that further numerical illustration of the bounds developed here, on a symmetric QEP arising from a finite element model of a simply supported beam, can be found in [8].

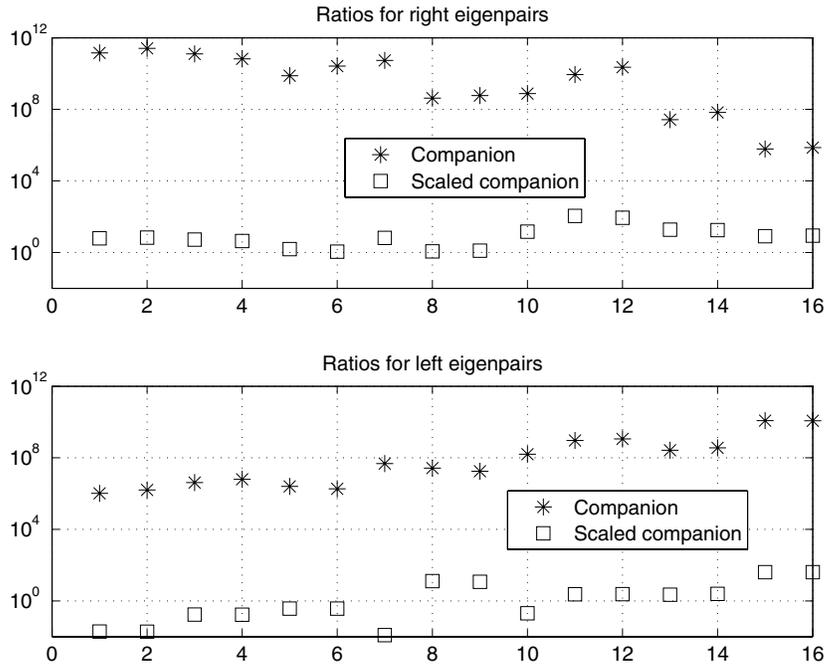


FIG. 6.1. Nuclear problem. Ratios η_Q/η_L for companion linearization $L = C_1$ and scaled companion linearization $L = D_s C_1$ for right eigenpairs (top) and left eigenpairs (bottom).

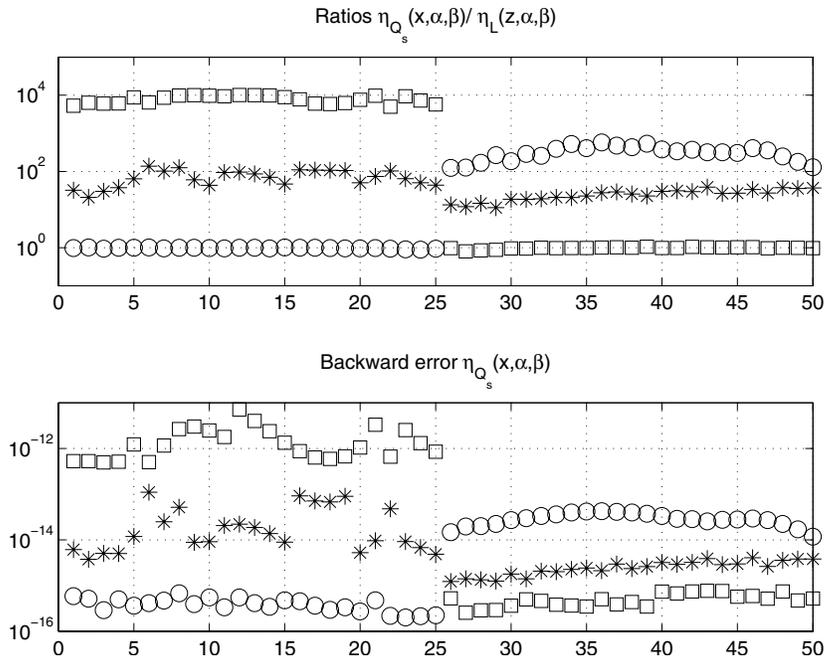


FIG. 6.2. Damped mass-spring problem. Ratios $\eta_{Q_s}(x, \alpha, \beta)/\eta_L(z, \alpha, \beta)$ and actual backward errors $\eta_{Q_s}(x, \alpha, \beta)$ with $x = z_1$ if $|\alpha| \geq |\beta|$ and $x = z_2$ otherwise, for $L = D_s C_1$ (*) and for $L = L_1$ (\square) and $L = L_2$ (\circ). Here, Q_s denotes the scaled quadratic \tilde{Q} .

TABLE 6.4
Damped mass-spring problem, $n = 50$.

Linearization L	Unscaled, $\rho = 3e2$		Scaled, $\rho = 1e2$	
	$\max \frac{\eta_Q}{\eta_L}$	Upper bound	$\max \frac{\eta_Q}{\eta_L}$	Upper bound
C_1	8.8e3	5.8e4	1.7e2	8.1e2
$D_s C_1$	1.0e2	9.0e2	1.4e2	4.1e2
L_1	2.0e3	2.9e4	1.0e4	1.5e5
L_2	1.8e3	1.1e5	5.7e2	3.8e4

Acknowledgment. We thank Steve Mackey for helpful discussions regarding the proof of Theorem 3.2.

REFERENCES

- [1] Z. BAI, J. W. DEMMEL, J. J. DONGARRA, A. RUHE, AND H. A. VAN DER VORST, EDs., *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, Software Environ. Tools 11, SIAM, Philadelphia, 2000.
- [2] H.-Y. FAN, W.-W. LIN, AND P. VAN DOOREN, *Normwise scaling of second order polynomial matrices*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 252–256.
- [3] P. FREITAS, M. GRINFELD, AND P. KNIGHT, *Stability of finite-dimensional systems with indefinite damping*, Adv. Math. Sci. Appl., 17 (1997), pp. 435–446.
- [4] I. GOHBERG, M. A. KAASHOEK, AND P. LANCASTER, *General theory of regular matrix polynomials and band Toeplitz operators*, Integral Equations Operator Theory, 11 (1988), pp. 776–882.
- [5] I. GOHBERG, P. LANCASTER, AND L. RODMAN, *Matrix Polynomials*, Academic Press, New York, 1982.
- [6] N. J. HIGHAM, D. S. MACKEY, N. MACKEY, AND F. TISSEUR, *Symmetric linearizations for matrix polynomials*, SIAM J. Matrix Anal. Appl., 29 (2006), pp. 143–159.
- [7] N. J. HIGHAM, D. S. MACKEY, AND F. TISSEUR, *The conditioning of linearizations of matrix polynomials*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1005–1028.
- [8] N. J. HIGHAM, D. S. MACKEY, F. TISSEUR, AND S. D. GARVEY, *Scaling, sensitivity and stability in the numerical solution of quadratic eigenvalue problems*, MIMS EPrint 2006.406, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2006. Internat. J. Numer. Methods Engrg., to appear.
- [9] N. J. HIGHAM, F. TISSEUR, AND P. M. VAN DOOREN, *Detecting a definite Hermitian pair and a hyperbolic or elliptic quadratic eigenvalue problem, and associated nearness problems*, Linear Algebra Appl., 351–352 (2002), pp. 455–474.
- [10] A. HILLIGES, C. MEHL, AND V. MEHRMANN, *On the solution of palindromic eigenvalue problems*, in Proceedings of the European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2004), Jyväskylä, Finland, P. Neittaanmäki, T. Rossi, S. Korotov, E. Oñate, J. Périaux, and D. Knörzer, eds., 2004. Available online from <http://www.mit.jyu.fi/eccomas2004/proceedings/proceed.html>.
- [11] I. C. F. IPSEN, *Accurate eigenvalues for fast trains*, SIAM News, 37 (2004), pp. 1–2.
- [12] T. ITOH, *Damped vibration mode superposition method for dynamic response analysis*, Earthquake Engrg. Struct. Dyn., 2 (1973), pp. 47–57.
- [13] P. LANCASTER, *Quadratic eigenvalue problems*, Linear Algebra Appl., 150 (1991), pp. 499–506.
- [14] P. LANCASTER AND P. PSARRAKOS, *A Note on Weak and Strong Linearizations of Regular Matrix Polynomials*, MIMS EPrint 2006.72, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2006.
- [15] P. LANCASTER AND M. TISMENETSKY, *The Theory of Matrices*, 2nd ed., Academic Press, London, 1985.
- [16] D. LEMONNIER AND P. M. VAN DOOREN, *Balancing regular matrix pencils*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 253–263.
- [17] D. S. MACKEY, *Structured Linearizations for Matrix Polynomials*, Ph.D. thesis, University of Manchester, Manchester, UK, 2006.
- [18] D. S. MACKEY, N. MACKEY, C. MEHL, AND V. MEHRMANN, *Structured polynomial eigenvalue problems: Good vibrations from good linearizations*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1029–1051.

- [19] D. S. MACKEY, N. MACKEY, C. MEHL, AND V. MEHRMANN, *Vector spaces of linearizations for matrix polynomials*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 971–1004.
- [20] C. B. MOLER AND G. W. STEWART, *An algorithm for generalized matrix eigenvalue problems*, SIAM J. Numer. Anal., 10 (1973), pp. 241–256.
- [21] G. W. STEWART, *On a companion operator for analytic functions*, Numer. Math., 18 (1971), pp. 26–43.
- [22] F. TISSEUR, *Backward error and condition of polynomial eigenvalue problems*, Linear Algebra Appl., 309 (2000), pp. 339–361.
- [23] F. TISSEUR AND K. MEERBERGEN, *The quadratic eigenvalue problem*, SIAM Rev., 43 (2001), pp. 235–286.
- [24] P. M. VAN DOOREN AND P. DEWILDE, *The eigenstructure of an arbitrary polynomial matrix: Computational aspects*, Linear Algebra Appl., 50 (1983), pp. 545–579.
- [25] R. C. WARD, *Balancing the generalized eigenvalue problem*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 141–152.