BACKWARD ERROR ANALYSIS OF POLYNOMIAL EIGENVALUE PROBLEMS SOLVED BY LINEARIZATION[*]

PIERS W. LAWRENCE[†], MARC VAN BAREL[‡], AND PAUL VAN DOOREN[†]

Abstract. We perform a backward error analysis of polynomial eigenvalue problems solved via linearization. Through the use of dual minimal bases, we unify the construction of strong linearizations for many different polynomial bases. By inspecting the prototypical linearizations for polynomials expressed in a number of classical bases, we are able to identify a small number of driving factors involved in the growth of the backward error. One of the primary factors is found to be the norm of the block vector of coefficients of the polynomial, which is consistent with the current literature. We derive upper bounds for the backward errors for specific linearizations, and these are shown to be reasonable estimates for the computed backward errors.

Key words. stability, backward error, polynomial eigenvalue problem, linearization, dual minimal basis, strong linearization

AMS subject classifications. 65F15, 15A18

DOI. 10.1137/15M1015777

1. Introduction. Polynomial matrices are everywhere, and their solution via linearization has been a central research topic for more than 50 years. Such problems are stated as solving the equation

$$(1.1) \qquad P(\lambda)x = 0 \, ,$$

where $P(\lambda)$ is a matrix of size $s \times s$ with entries that are polynomial in λ, and where x is a nonzero vector of length s. The pair (λ, x) is called a right eigenpair of the polynomial $P(\lambda)$. One is often also interested in left eigenpairs—pairs (λ, y^*) satisfying $y^*P(\lambda) = 0$—but here we concentrate mainly on right eigenpairs.

Our analysis considers three different polynomial bases in which we express the polynomial $P(\lambda)$ in (1.1): the monomial, Chebyshev, and Lagrange bases. For each of these bases, suppose that we are given a set of elements $\phi_i(\lambda)$, $0 \leq i \leq N$, forming a basis for $\mathcal{P}_N$—the set of polynomials of degree at most N. Any polynomial matrix $P(\lambda) \in \mathcal{P}_N$ can thus be expressed as

$$(1.2) \qquad P(\lambda) = \sum_{i=0}^{N} P_i \phi_i(\lambda) \, ,$$

where $P_i \in \mathbb{C}^{s \times s}$ are the coefficients in the specified basis.

[†]Department of Mathematical Engineering, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium (piers.lawrence@uclouvain.be, paul.vandooren@uclouvain.be).

[‡]Department of Computer Science, KU Leuven, B-3001 Leuven (Heverlee), Belgium (marc.vanbarel@cs.kuleuven.be).

Of the plethora of methodologies proposed for solving (1.1), one of the most common solution strategies is linearization. The polynomial problem is transformed into a linear eigenvalue problem that has both the same eigenvalues as $P(\lambda)$ and eigenvectors that are easily related. The linearization should be constructed from the coefficients P_i with little or no computation necessary.

2. Linearization and dual minimal bases. In this section, we revisit the notions of minimal polynomial bases and their minimal indices. We also recall the notion of dual minimal bases, which will play a crucial role in our constructive algorithms.

To begin with, let us review the definitions of *row reduced* and *minimal bases* for polynomial matrices. Kailath [15, Chap. 6] offers the following definitions.

DEFINITION 2.1. *An $m \times n$ polynomial matrix $N(\lambda)$ with the highest row degrees $\{d_1, \ldots, d_m\}$ is called* row reduced *if the highest degree coefficient matrix N_h (whose jth row is the coefficient of λ^{d_j} in the jth row of $N(\lambda)$) has full row rank.*

DEFINITION 2.2. *The rows of an $m \times n$ polynomial matrix $N(\lambda)$ are a* minimal basis *if $N(\lambda)$ has full row rank m for all finite $\lambda \in \mathbb{C}$ and is row reduced. The corresponding row degrees are called* minimal indices *of that basis.*

DEFINITION 2.3. *Let the rows of two polynomial matrices $N_1(\lambda)$ and $N_2(\lambda)$ be minimal bases of dimension $m_1 \times n$ and $m_2 \times n$, respectively. Such matrices are said to be dual minimal bases if $N_1(\lambda)N_2^T(\lambda) = 0$ and $m_2 := n - m_1$.*

Given these definitions, we present a lemma that defines the properties of a dual minimal basis, which we will later use to construct linearizations of polynomials.

LEMMA 2.4. *Let the rows of $N_1(\lambda)$ be a given minimal basis of dimension $m_1 \times n$ with minimal indices η_j, $j = 1, \ldots, m_1$. Every dual minimal basis $N_2(\lambda)$ thus has dimension $m_2 \times n$ with $m_2 := n - m_1$. Moreover, its minimal indices ϵ_i, $i = 1, \ldots, m_2$ add up to the same sum as those of $N_1(\lambda)$:*

$$\sum_{j=1}^{m_1} \eta_j = \sum_{i=1}^{m_2} \epsilon_i \,.$$

Proof. This result can be found in the corollary to Theorem 3 of Forney [9]; see also Kailath [15, sect. 6.5.4] for a more accessible discussion of the same result. □

In this section, we focus on the construction of strong linearizations—of a particular order, i.e., the dimension of the matrices—for a given polynomial matrix. We consider the case where the polynomial matrix is regular, in that it has nonidentically zero determinant. Further, we consider polynomials of *grade* $(N + 1)$, that is, of the form

$$(2.1) \qquad P(\lambda) = P_{N+1}\lambda^{N+1} + P_N\lambda^N + \cdots + P_1\lambda + P_0 \,,$$

where we allow the polynomial to have degree smaller than $(N + 1)$; in this work we will explicitly take $P_{N+1} = 0$. The introduction of such a leading coefficient gives the ability to construct linearizations where the polynomial coefficients are separated from the relations between the basis elements. According to Gohberg, Kaashoek, and Lancaster [10], a linearization of order $(N + 1)s$ of such a regular polynomial matrix is a regular pencil

$$(2.2) \qquad \mathcal{L}(\lambda) = \lambda B - A$$

of dimension $(N + 1)s$ (that is, its order), for which there exist unimodular transfor-

mations $M(\lambda)$ and $N(\lambda)$ such that

$$(2.3) \qquad M(\lambda)\mathscr{L}(\lambda)N(\lambda) = \left[\begin{array}{cc} P(\lambda) & 0 \\ 0 & I_{Ns} \end{array} \right].$$

A *strong* linearization of order $(N+1)s$ (defined in [10] and so named in [16]) is one where a similar relation also holds for the reversed polynomial

$$\mathrm{rev}P(\lambda) := P_0\lambda^{N+1} + \cdots + P_N\lambda + P_{N+1}$$

and pencil $\mathrm{rev}\mathscr{L}(\lambda) := B - \lambda A$. Note that throughout this manuscript, we take the reverse polynomial with respect to the *grade* of the polynomial; this is quite consistent with the original notion of strong linearization of a particular order [10, 17, 16].

Let us now give a technical lemma that will allow us to construct large families of strong linearizations of order $(N+1)s$.

LEMMA 2.5. *Let $P(\lambda)$ be a given $s \times s$ regular polynomial matrix. Let $\mathscr{L}(\lambda)$ be a regular pencil of the form*

$$(2.4) \qquad \mathscr{L}(\lambda) := \left[\begin{array}{c} L_1(\lambda) \\ L_2(\lambda) \end{array} \right] = \left[\begin{array}{c} -A_1 \\ \lambda B_2 - A_2 \end{array} \right],$$

where $L_1(\lambda)$ and $L_2(\lambda)$ have dimensions $s \times (N+1)s$ and $Ns \times (N+1)s$, respectively. Let $\Phi(\lambda)$ be another polynomial matrix

$$(2.5) \qquad \Phi(\lambda) := \Phi_0 + \Phi_1\lambda + \cdots + \Phi_N\lambda^N ,$$

of dimension $(N+1)s \times s$ and degree N, that satisfies

$$(2.6) \qquad \mathscr{L}(\lambda)\Phi(\lambda) = \left[\begin{array}{c} P(\lambda) \\ 0 \end{array} \right].$$

$\mathscr{L}(\lambda)$ *is thus a strong linearization of $P(\lambda)$ of order $(N+1)s$ if the following conditions hold:*

1. $L_2(\lambda)$ *has full row rank for all $\lambda \in \mathbb{C}$;*
2. B_2 *has full row rank;*
3. $\Phi(\lambda)$ *has full column rank for all $\lambda \in \mathbb{C}$; and*
4. Φ_N *has full column rank.*

Proof. The conditions on $L_2(\lambda)$ and $\Phi(\lambda)$ mean that the rows of $L_2(\lambda)$ are a minimal polynomial basis for the row space of $L_2(\lambda)$ and that the columns of $\Phi(\lambda)$ are a minimal polynomial basis for the column space of $\Phi(\lambda)$. Since $L_2(\lambda)\Phi(\lambda) = 0$ and the row degrees of $L_2(\lambda)$ and column degrees of $\Phi(\lambda)$ both add up to Ns, these two bases are also dual. De Terán, Dopico, and Van Dooren [6, Thm. 4.1] have shown that, under these conditions, there exist polynomial matrices $X(\lambda)$ and $Y(\lambda)$ such that the matrix $\left[\begin{array}{c|c} \Phi(\lambda) & Y(\lambda) \end{array} \right]$ is unimodular, and that

$$(2.7) \qquad \left[\begin{array}{cc} I_s & X(\lambda) \\ 0 & I_{Ns} \end{array} \right] \left[\begin{array}{c} L_1(\lambda) \\ \hline L_2(\lambda) \end{array} \right] \left[\begin{array}{c|c} \Phi(\lambda) & Y(\lambda) \end{array} \right] = \left[\begin{array}{cc} P(\lambda) & 0 \\ 0 & I_{Ns} \end{array} \right].$$

This implies that $\mathscr{L}(\lambda)$ is a linearization of order $(N+1)s$ of $P(\lambda)$. However, the

conditions on $L_2(\lambda)$ and $\Phi(\lambda)$ are reversible, implying that they also hold for $\mathrm{rev}L_2(\lambda)$ and $\mathrm{rev}\Phi(\lambda)$. The rows of $\mathrm{rev}L_2(\lambda)$ and the columns of $\mathrm{rev}\Phi(\lambda)$ are dual, that is,

$$(2.8) \qquad (B_2 - \lambda A_2)(\Phi_0 \lambda^N + \cdots + \Phi_{N-1}\lambda + \Phi_N) = 0 \,,$$

and are minimal polynomial bases. Thus, there exist left and right unimodular transformations, such that

$$(2.9) \qquad \begin{bmatrix} I_s & \widehat{X}(\lambda) \\ 0 & I_{Ns} \end{bmatrix} \begin{bmatrix} \mathrm{rev}L_1(\lambda) \\ \hline \mathrm{rev}L_2(\lambda) \end{bmatrix} \begin{bmatrix} \mathrm{rev}\Phi(\lambda) & | & \widehat{Y}(\lambda) \end{bmatrix} = \begin{bmatrix} \mathrm{rev}P(\lambda) & 0 \\ 0 & I_{Ns} \end{bmatrix} .$$

Therefore, $\mathscr{L}(\lambda)$ is also a *strong* linearization of order $(N+1)s$ of $P(\lambda)$.   □

Remark 1.   One can find some similarities between the approach taken in Lemma 2.5 and the work of Grammont, Higham, and Tisseur [12] to the extent that they, too, construct what they call one-sided factorizations of the linearization in the form (2.6). In that work they use the factorization to relate specific eigenpairs of the linearization to those of the polynomial. The relationships can be used to obtain *local* relationships between the backward errors of *specific eigenpairs* of the linearization to the eigenpairs of the polynomial. In the present manuscript, we show in section 5 how we can obtain *global* backward error bounds for the polynomial coefficients.

The above Lemma does not guarantee that every polynomial matrix $P(\lambda)$ can be linearized in this way. In order to do that there must be a solution A_1 to the equation

$$(2.10) \qquad -A_1 \begin{bmatrix} \Phi_N & \cdots & \Phi_0 \end{bmatrix} = \begin{bmatrix} P_N & \cdots & P_0 \end{bmatrix}$$

(that is, to the top block of (2.6)) for every $P(\lambda)$ of grade $N+1$ with $P_{N+1} = 0$. This will be the case if and only if the $(N+1)s \times (N+1)s$ matrix

$$(2.11) \qquad \Phi := \begin{bmatrix} \Phi_N & \cdots & \Phi_0 \end{bmatrix}$$

is invertible. This is indeed the case since the left null vectors $y(\lambda)^T \Phi(\lambda) = 0$ belong to the row space of $\lambda B_2 - A_2$ and B_2 has full row rank. Otherwise, there would be a constant vector that is dual to the column space of $\Phi(\lambda)$, contradicting the properties of minimal bases and the assumptions of Lemma 2.5 (that is, that B_2 has full row rank). We thus state the following theorem concerning the construction of such strong linearizations.

THEOREM 2.6. *Let $L_2(\lambda) = \lambda B_2 - A_2$ of dimension $Ns \times (N+1)s$ and $\Phi(\lambda)$ of dimension $(N+1)s \times s$ be dual minimal bases with row degrees $\eta_j = 1$, $1 \le j \le Ns$, and column degrees $\epsilon_i = N$, $1 \le i \le N$, respectively. Thus, there is a solution, A_1, to the equation*

$$(2.12) \qquad \begin{bmatrix} -A_1 \\ \lambda B_2 - A_2 \end{bmatrix} \Phi(\lambda) = \begin{bmatrix} P(\lambda) \\ 0 \end{bmatrix}$$

for every polynomial matrix $P(\lambda)$ of grade $N+1$ with $P_{N+1} = 0$, and $\mathscr{L}(\lambda) = \begin{bmatrix} -A_1 \\ \lambda B_2 - A_2 \end{bmatrix}$ is a strong linearization of order $(N+1)s$ of $P(\lambda)$.

Proof. The combination of Lemma 2.5 and the condition that the matrix Φ defined in (2.11) is invertible guarantees the existence of such a solution.   □

COROLLARY 2.7. *The duality equation $L_2(\lambda)\Phi(\lambda) = 0$ corresponds to the equation*

$$(2.13) \qquad C(A_2, B_2)\Psi = 0 \,,$$

where $C(A_2, B_2)$ is a matrix of dimension $(N+2)Ns \times (N+1)^2 s$ that expresses the coefficients of $L_2(\lambda)\Phi(\lambda)$ in the same basis as $\Phi(\lambda)$, with the addition of one extra element of degree $(N+1)$. For example, in the monomial basis, these are defined as

$$(2.14) \qquad C(A_2, B_2) := \begin{bmatrix} B_2 & & & \\ -A_2 & \ddots & & \\ & \ddots & B_2 \\ & & -A_2 \end{bmatrix}, \text{ and } \Psi := \begin{bmatrix} \Phi_N \\ \vdots \\ \Phi_0 \end{bmatrix}.$$

Remark 2. The rank conditions defined in Lemma 2.5 on $L_2(\lambda)$ and $\Phi(\lambda)$ imply that the matrix $C(A_2, B_2)$ has full row rank $(N+2)Ns$ and that Ψ is a full column rank matrix spanning the s-dimensional null space of $C(A_2, B_2)$. Since the full rank properties also hold in an open neighborhood of the given matrices, this guarantees that we can talk about the Taylor expansion of this matrix equation when considering small perturbations. A first-order perturbation analysis therefore makes sense, and it gives a smooth parameterization of the perturbed polynomial matrix derived from this expansion. It also guarantees that the perturbed pencil corresponds to a strong linearization of that perturbed polynomial matrix.

3. Classes of linearizations. In this work, we restrict our attention to linearizations of polynomials expressed in three specific bases: the monomial, Chebyshev, and Lagrange bases. These three categories are those most commonly encountered in practice, many different linearizations exist for them, and they represent the prototypical forms for more general polynomial bases.

Throughout, we will discuss linearizations of order $(N+1)s$, that is, having the form (2.4), and thus we depart from many other works that only consider linearizations of order Ns. We use this seemingly trivial device of adding an additional s eigenvalues at infinity in order to unify the analysis of all three polynomial bases that we consider.

By using linearizations of the form (2.4), we are able to construct and analyze linearizations where the relations between the basis elements are separated from the polynomial coefficients. Furthermore, for many specific linearizations of interest, we can often organize the computations in a simple way to decouple the spurious infinite eigenvalues via constant equivalence transformations applied to the linearization.

For each of the three polynomial bases that we consider in this manuscript, we specialize the linearization (2.4) to have $-A_1 = \begin{bmatrix} P_N & \dots & P_0 \end{bmatrix}$, where the P_i's are the polynomial coefficients in that particular basis. For each basis, we also specify the linearization $\mathscr{L}(\lambda)$ and a matrix $\Phi(\lambda)$ together satisfying (2.6).

3.1. Monomial basis. One of the most well-known linearizations is the companion pencil. Of its many incarnations, the one we use in this work is a generalization—but a straightforward one—defined as

$$(3.1) \qquad \mathscr{L}(\lambda) = \begin{bmatrix} P_N & P_{N-1} & \dots & P_0 \\ -I & \lambda I & & \\ & \ddots & \ddots & \\ & & -I & \lambda I \end{bmatrix}, \quad \Phi(\lambda) = \begin{bmatrix} \lambda^N I \\ \vdots \\ \lambda I \\ I \end{bmatrix},$$

where the P_i's are the monomial basis coefficients.

3.2. Chebyshev basis. Good [11] developed one of the first linearizations for scalar polynomials expressed in the Chebyshev basis, naming them *colleague* matrices. Although Good did not consider matrix polynomials, the generalization of the

colleague matrix is straightforward and can be given by

$$
(3.2) \quad \mathscr{L}(\lambda) = \begin{bmatrix} P_N & P_{N-1} & \cdots & P_1 & P_0 \\ -\frac{1}{2}I & \lambda I & -\frac{1}{2}I & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{2}I & \lambda I & -\frac{1}{2}I \\ & & & -I & \lambda I \end{bmatrix}, \quad \Phi(\lambda) = \begin{bmatrix} T_N(\lambda)I \\ T_{N-1}(\lambda)I \\ \vdots \\ T_1(\lambda)I \\ T_0(\lambda)I \end{bmatrix},
$$

where $T_i(\lambda)$ is the ith Chebyshev polynomial of the first kind.

3.3. Lagrange basis. A number of works have considered linearization of polynomials expressed in the Lagrange basis. For the scalar case, Fiedler [8] examined the construction of symmetric matrices with a given characteristic polynomial and produced linearizations of arrowhead form. Malek and Vaillancourt [20] developed a method of Newbery [22] to a method for computing a matrix with a given characteristic polynomial that results in a construction similar to Fiedler's linearization.

Following Berrut and Trefethen's [3] recent rediscovery of the *barycentric* formula for the Lagrange interpolating polynomial, there has been a renewed interest in the use of Lagrange interpolation, spurring on the development of rootfinding methods specifically designed for the Lagrange basis. The barycentric Lagrange interpolation formula is given as

$$
(3.3) \quad P(\lambda) = \ell(\lambda) \sum_{i=0}^{N} \frac{\beta_i P_i}{\lambda - \sigma_i},
$$

where the σ_i's are distinct interpolation nodes, and where the node polynomial, barycentric weights, and Lagrange polynomials are defined by

$$
(3.4) \quad \ell(\lambda) = \prod_{i=0}^{N} (\lambda - \sigma_i), \quad \beta_i^{-1} = \prod_{\substack{j=0 \\ j \neq i}}^{N} (\sigma_i - \sigma_j), \quad \text{and} \quad \ell_i(\lambda) = \ell(\lambda) \frac{\beta_i}{\lambda - \sigma_i},
$$

respectively. The barycentric Lagrange formula was developed for the scalar interpolation problem. However, if the P_i's are matrices, then the formula still interpolates the polynomial matrix with values $P_i = P(\sigma_i)$; thus, the formula is applicable to polynomial matrices with no further modification.

For matrix polynomials expressed in barycentric Lagrange form, one of the first linearizations was proposed by Corless [4]. Amiraslani, Corless, and Lancaster [1] also later developed linearizations for polynomials in other bases.

A defining property of the Lagrange basis is that all of the basis elements $\ell_i(\lambda)$ are of degree N, and hence—unlike the monomial and Chebyshev bases—there is much greater choice for the relations between the basis elements used to construct the linearization. We examine two linearizations that are modifications of those proposed by Corless [4] and by Van Beeumen, Michiels, and Meerbergen [27]. The first,

$$
(3.5) \quad \mathscr{L}(\lambda) = \begin{bmatrix} 0 & P_N & \cdots & P_0 \\ \beta_N I & (\lambda - \sigma_N)I & & \\ \vdots & & \ddots & \\ \beta_0 I & & & (\lambda - \sigma_0)I \end{bmatrix}, \quad \Phi(\lambda) = \begin{bmatrix} \ell(\lambda)I \\ \ell_N(\lambda)I \\ \vdots \\ \ell_0(\lambda)I \end{bmatrix},
$$

is just a rearrangement of the linearization proposed by Corless [4]. In our numerical experience—using the QZ algorithm to compute the eigenvalues—we have found that

the spurious infinite eigenvalues are better deflated from this arrangement of the linearization.

The second linearization, proposed by Van Beeumen, Michiels, and Meerbergen [27], is given by

(3.6)

$$\mathscr{L}(\lambda) = \begin{bmatrix} P_N & P_{N-1} & \cdots & P_1 & P_0 \\ (\lambda - \sigma_N)I & -(\lambda - \sigma_{N-1})\theta_{N-1}I & & & \\ & (\lambda - \sigma_{N-1})I & \ddots & & \\ & & \ddots & -(\lambda - \sigma_1)\theta_1 I & \\ & & & (\lambda - \sigma_1)I & -(\lambda - \sigma_0)\theta_0 I \end{bmatrix},$$

(3.7)

$$\Phi(\lambda) = \begin{bmatrix} \ell_N(\lambda)I \\ \ell_{N-1}(\lambda)I \\ \vdots \\ \ell_1(\lambda)I \\ \ell_0(\lambda)I \end{bmatrix},$$

where $\theta_i = \beta_{i+1}/\beta_i$. The linearization (3.5) does not precisely fit into the framework described in the preceding section, since it is a linearization of order $(N+2)s$. However, it is worth stating that the second linearization can be obtained by multiplying $\mathscr{L}(\lambda)$ on the left by a sequence of Gauss transformations $G_{N-1} \cdots G_0$, where

(3.8)

$$G_i = \left[\begin{array}{c|cc|c} I_{(N-i)s} & & & \\ \hline & I & 0 & \\ & I & -\theta_i I & \\ \hline & & & I_{is} \end{array} \right],$$

are chosen to annihilate the entries of the first block column from the bottom to the top of the matrix. Following these transformations, we may deflate s infinite eigenvalues by permuting the first two block rows to obtain the smaller linearization (3.6). This procedure to deflate infinite eigenvalues from the pencil (3.5) has been recently described by Lawrence and Corless [19], where they also describe how to deflate other infinite eigenvalues from the pencil whenever P_N is singular.

4. Backward errors. In this section, we review the concept of backward errors for the solutions of polynomial eigenvalue problems solved via linearization. Suppose we have computed an approximate solution of (1.1), that is, we have a computed value $\widehat{\lambda} \in \mathbb{C}$ and a vector $\widehat{x} \in \mathbb{C}^s$ that are approximations to the true solution of (1.1). A natural question to ask is how close this solution is to the true solution; this distance is known as the forward error. Another natural question to ask is this: what problem has actually been solved? By this we mean that given a computed eigenpair $(\widehat{\lambda}, \widehat{x})$, for what perturbation matrix $\Delta P(\lambda)$ is

(4.1)
$$(P(\widehat{\lambda}) + \Delta P(\widehat{\lambda}))\widehat{x} = 0$$

satisfied? If the norm of the coefficients of $\Delta P(\lambda)$ are of the same order as the error in the input data, then one can hardly criticize such a solution $(\widehat{\lambda}, \widehat{x})$. It is the exact solution to a nearby problem, where nearby can be viewed as being of the size of the errors in the input data.

Tisseur [25] defines the normwise backward error of an approximate eigenpair as

$$(4.2) \quad \eta_P(\widehat{\lambda}, \widehat{x}) := \min \left\{ \varepsilon : (P(\widehat{\lambda}) + \Delta P(\widehat{\lambda}))\widehat{x} = 0, \ \|\Delta P_i\|_2 \le \varepsilon \|P_i\|_2, \ 0 \le i \le N \right\},$$

where ΔP_i and P_i are the monomial basis coefficients of $\Delta P(\lambda)$ and $P(\lambda)$, respectively. Furthermore, Tisseur shows that the following simple formula can be used to compute the backward error of an approximate eigenpair:

$$(4.3) \qquad \eta_P(\widehat{\lambda}, \widehat{x}) = \frac{\|P(\widehat{\lambda})\widehat{x}\|_2}{\left(\sum_{i=0}^{N} \|P_i\|_2 |\widehat{\lambda}|^i \right) \|\widehat{x}\|_2}.$$

Another useful expression for the backward errors of computed eigenvalues was also given by Tisseur [25, Lem. 3]; this expression is independent of the specific computed eigenvectors given by

$$(4.4) \qquad \eta_P(\widehat{\lambda}) = \frac{\|P(\widehat{\lambda})^{-1}\|_2^{-1}}{\sum_{i=0}^{N} \|P_i\|_2 |\widehat{\lambda}|^i}.$$

5. Perturbation theory. In this section, we examine the backward errors of solutions to polynomial eigenvalue problems found via linearization. To do so, we examine how generic perturbations can—to first order—be mapped onto structured perturbations of the polynomial coefficients. The techniques described here are inspired by the works of Van Dooren and Dewilde [28], and also of Edelman and Murakami [7], where they examine such mappings. In contrast to those works, we are able to derive upper bounds for the perturbations to the polynomial coefficients obtained from these mappings. Let us comment on some of the other approaches currently taken in the literature. There have been many articles performing backward error analyses of scalar polynomial rootfinding via linearization. These can be summarized in the works [7, 18, 21] and the more recent works [5, 23]. For matrix polynomials, there have been a number of backward error analyses performed, for example, Higham, Li, and Tisseur [13] give local bounds for specific eigenpairs of the companion linearizations, as do Lawrence and Corless [19] for linearizations in the Lagrange basis. Van Dooren and Dewilde [28] perform a global analysis for the companion linearization where they obtain larger upper bounds for the backward errors.

Let us consider the following procedure for computing a solution to the polynomial eigenvalue problem. Suppose we are given the coefficients of a polynomial matrix in a particular basis. We construct the appropriate linearization and then pass the matrices off to the QZ algorithm in order to compute the eigenvalues. According to the LAPACK Users' Guide [2, sect. 4.11.1], the QZ algorithm is normwise backward stable. In other words, the algorithm computes the exact eigenvalues, eigenvectors, and deflating subspaces of a slightly perturbed pencil,

$$(5.1) \qquad\qquad \lambda(B + \Delta B) - (A + \Delta A),$$

where the perturbations satisfy $\|(\Delta A, \Delta B)\|_F \le q(Ns)\varepsilon_M \|(A, B)\|_F$, the Frobenius norm of a pencil is defined as $\|(A, B)\|_F = \sqrt{\|A\|_2^2 + \|B\|_2^2}$, $q(Ns)$ is a polynomial function of the dimension of the linearization (usually taken to be linear in practice), and ε_M is the machine epsilon.

This backward stability, of course, only applies to the solution of the linear generalized eigenvalue problem. What we would really like to know is if we have backward

stability with respect to the polynomial problem: what polynomial eigenvalue problem have we really solved, and is the distance to the original polynomial small?

We will perform a first-order backward error analysis by simultaneously considering the perturbation of the linearization $\mathscr{L}(\lambda)$ and the one-sided factorization $\Phi(\lambda)$ in order to retain the structure (2.12) for the perturbed problem. More concretely, we ensure—to first order—that the perturbed problem has the following structure:

$$(5.2) \qquad \left[\begin{array}{c} \lambda \Delta B_1 - (A_1 + \Delta A_1) \\ \lambda(B_2 + \Delta B_2) - (A_2 + \Delta A_2) \end{array} \right] (\Phi(\lambda) + \Delta\Phi(\lambda)) = \left[\begin{array}{c} P(\lambda) + \Delta P(\lambda) \\ 0 \end{array} \right].$$

In other words, ensuring that $\lambda(B_2 + \Delta B_2) - (A_2 + \Delta A_2)$ and $\Phi(\lambda) + \Delta\Phi(\lambda)$ are dual minimal bases, we are able to map generic perturbations of the linearization onto structured perturbations of the polynomial coefficients. A good backward error result is that the norms of the coefficients of $\Delta P(\lambda)$ are relatively small. Once we compute a first-order solution $\Delta\Phi(\lambda)$ that satisfies (5.2), we recover the polynomial coefficients as

$$(5.3) \qquad \Delta P(\lambda) = (\lambda \Delta B_1 - \Delta A_1)\Phi(\lambda) - A_1 \Delta\Phi(\lambda).$$

The first term in this expansion is of the same order of the perturbations of the linearization, that is, $O(\varepsilon_M)$. Thus, in order to have good backward errors, it seems appropriate to compute a solution $\Delta\Phi(\lambda)$ that minimizes the contribution from the second term. To compute such a solution for the perturbations to the one-sided factorization, consider the bottom block of (5.2). To first order, we have

$$(5.4) \qquad (\lambda B_2 - A_2)\Delta\Phi(\lambda) = -(\lambda \Delta B_2 - \Delta A_2)\Phi(\lambda).$$

By equating coefficients—in the specific polynomial basis in which $\Phi(\lambda)$ is expressed—we obtain a linear system for the coefficients of $\Delta\Phi(\lambda)$. The coefficients of $\Delta\Phi(\lambda)$ satisfy

$$(5.5) \qquad C(A_2, B_2)\Delta\Psi = -C(\Delta A_2, \Delta B_2)\Psi,$$

where

$$(5.6) \qquad \Delta\Psi := \left[\begin{array}{c} \Delta\Phi_N \\ \vdots \\ \Delta\Phi_0 \end{array} \right] \text{ and } \Psi := \left[\begin{array}{c} \Phi_N \\ \vdots \\ \Phi_0 \end{array} \right]$$

are block vectors of the coefficients $\Delta\Phi(\lambda)$ and $\Phi(\lambda)$, respectively, and where $C(A, B)$ is a matrix constructed from the matrices A and B, and for the specific polynomial basis at hand (see, for example, the matrix in (2.14)). We compute a minimum norm solution to (5.5), and this seems to be a reasonable thing to do, since the growth in (5.3) is associated with $\Delta\Phi(\lambda)$. A minimum norm solution in any unitarily invariant norm is given by

$$(5.7) \qquad \Delta\Psi = -C(A_2, B_2)^+ C(\Delta A_2, \Delta B_2)\Psi,$$

where $C(A_2, B_2)^+$ is the Moore–Penrose pseudoinverse of $C(A_2, B_2)$. We inspect the polynomial coefficients of $\Delta P(\lambda)$ given in (5.3), and together with the matrices $C(A_2, B_2)$ for the specific polynomial basis, we obtain

$$(5.8) \qquad \left[\begin{array}{c} \Delta P_{N+1} \\ \Delta P_N \\ \vdots \\ \Delta P_0 \end{array} \right] =: \overrightarrow{\Delta P} = C(\Delta A_1, \Delta B_1)\Psi - C(A_1, 0)C(A_2, B_2)^+ C(\Delta A_2, \Delta B_2)\Psi.$$

The above equation holds for any strong linearization of the form (2.4) given in Theorem 2.6. In the next section, we will obtain bounds for the norms of these coefficients for the specific linearizations and bases that we introduced in section 3. To show the primary factors involved in the growth of the coefficients of the perturbations, we offer the following lemma.

LEMMA 5.1. *The block vector $\overrightarrow{\Delta P}$ of coefficients of the perturbation $\Delta P(\lambda)$ can be bounded by*

$$(5.9) \quad \left\|\overrightarrow{\Delta P}\right\|_2 \le \|C(\Delta A_1, \Delta B_1)\Psi\|_2 + \|A_1\|_2 \|C(A_2, B_2)^+\|_2 \|C(\Delta A_2, \Delta B_2)\Psi\|_2 \,,$$

where Ψ is the block vector of the coefficients of $\Phi(\lambda)$ defined in (5.6).

Proof. Simply take the norm of the block coefficient vector (5.8). We have $B_{N+1} = [\ 0 \quad I_{N+1}\ ]$, so that $C(A_1, 0)$ is block diagonal, and hence $\|C(A_1, 0)\|_2 = \|A_1\|_2$, and the upper bound follows. $\quad\square$

Remark 3. We see that there is the possibility of perturbing the coefficient P_{N+1}, and thus s eigenvalues could in theory be perturbed to become large finite eigenvalues. When applying the QZ algorithm to the pencils, such perturbations can be avoided; by a preliminary deflation of these s eigenvalues, we can ensure that they remain at infinity. One will also note that this condition arises only in the first term of (5.8), and for the linearizations considered in this manuscript $\Phi_N = e_1 \otimes I_s$. We thus see that the condition $\Delta P_{N+1} = \Delta B_1(e_1 \otimes I_s) = 0$ requires that we do not perturb the principal $s \times s$ submatrix of B.

6. Growth of $\|C(A_2, B_2)^+\|_2$. In this section we consider the growth of the upper bound (5.9) determined by the term $\|C(A_2, B_2)^+\|_2$. This term and the norm of the coefficient vector $\|A_1\|_2$ are the primary driving factors determining the growth of the upper bound. Based on the explicit matrices $C(A_2, B_2)$ for the specific basis considered, we determine bounds for the norm $\|C(A_2, B_2)^+\|_2$. We consider each of the linearizations given in section 3: (3.1) for the monomial basis, (3.2) for the Chebyshev basis, and (3.6) for the Lagrange basis.

We first make the observation that all of the linearizations have a Kronecker product construction, by which we mean that we are able to express $A_2 = \widehat{A}_N \otimes I_s$ and $B_2 = \widehat{B}_N \otimes I_s$ for each linearization. Furthermore, this means that each matrix $C(A_2, B_2)$ can be written in the form

$$(6.1) \quad C(A_2, B_2) = (\widehat{A}_{N+1}^T \otimes \widehat{B}_N - \widehat{B}_{N+1}^T \otimes \widehat{A}_N) \otimes I_s \,,$$

where $\widehat{A}_N$ and $\widehat{B}_N$ are directly related to the linearization matrices A_2 and B_2, respectively, and where $\widehat{A}_{N+1}$ and $\widehat{B}_{N+1}$ are closely related matrices derived from the relations between the elements of the polynomial basis. In all cases, we define the explicit matrices for each of the polynomial bases involved.

To compute bounds for the backward errors described in (5.9), we should compute bounds for the norm $\|C(A_2, B_2)^+\|_2 = 1/\sigma_{\min}(C(A_2, B_2))$, where $\sigma_{\min}(C(A_2, B_2))$ is the smallest singular value of $C(A_2, B_2)$. Thus, with the aid of (6.1) we see that the computation reduces to

$$(6.2) \quad \|C(A_2, B_2)^+\|_2 = \frac{1}{\sigma_{\min}(\widehat{A}_{N+1}^T \otimes \widehat{B}_N - \widehat{B}_{N+1}^T \otimes \widehat{A}_N)} \,.$$

6.1. Monomial. The construction of the matrix $C(A_2, B_2)$ for the monomial basis is perhaps the easiest to identify. Consider the expansion of the product

$$(6.3) \qquad (\lambda B_2 - A_2)\Phi(\lambda) = \Theta\lambda^{N+1} + \sum_{i=0}^{N}\Theta_i\lambda^i \,,$$

that is, the product expressed in the same monomial basis with the addition of the extra basis element λ^{N+1}. This equation can be written as the linear system

$$(6.4) \qquad \begin{bmatrix} \Theta \\ \Theta_N \\ \vdots \\ \Theta_1 \\ \Theta_0 \end{bmatrix} = \begin{bmatrix} B_2 & & & \\ -A_2 & B_2 & & \\ & \ddots & \ddots & \\ & & -A_2 & B_2 \\ & & & -A_2 \end{bmatrix} \begin{bmatrix} \Phi_N \\ \Phi_{N-1} \\ \vdots \\ \Phi_1 \\ \Phi_0 \end{bmatrix}.$$

Inspecting the linearization given in (3.1), we see that $A_2 = \widehat{A}_N \otimes I_s$ and $B_2 = \widehat{B}_N \otimes I_s$, where $\widehat{A}_N$ and $\widehat{B}_N$ are given by

$$(6.5) \qquad \widehat{A}_N = \begin{bmatrix} I_N & 0 \end{bmatrix}, \quad \widehat{B}_N = \begin{bmatrix} 0 & I_N \end{bmatrix}.$$

Furthermore, the matrix $C(A_2, B_2)$ in (6.4) can be rewritten in the form (6.1), where

$$(6.6) \qquad \widehat{A}_{N+1} = \begin{bmatrix} I_{N+1} & 0 \end{bmatrix}, \quad \widehat{B}_{N+1} = \begin{bmatrix} 0 & I_{N+1} \end{bmatrix}.$$

Now we will give upper and lower bounds for the norm of the pseudoinverse in the following lemma.

LEMMA 6.1. *For the monomial basis linearization* (3.1) *the norm* $\|C(A_2, B_2)^+\|_2$ *satisfies the following bounds:*

$$(6.7) \qquad \frac{1}{\sqrt{2\left(1 + \cos\left(\frac{\pi}{2}\frac{2N-1}{N}\right)\right)}} \leq \left\|C(A_2, B_2)^+\right\|_2 \leq \frac{1}{\sqrt{2\left(1 + \cos\left(\frac{\pi}{2}\frac{2N+1}{N+1}\right)\right)}}.$$

Proof. As discussed in the introduction to this section, the matrix $C(A_2, B_2)$ for the monomial basis (6.4) can be rewritten in the form (6.1), where the matrices $\widehat{A}_N$ and $\widehat{B}_N$ are defined in (6.5), and where $\widehat{A}_{N+1}$ and $\widehat{B}_{N+1}$ are defined in (6.6). It can be seen that the matrix $\widehat{A}_{N+1}^T \otimes \widehat{B}_N - \widehat{B}_{N+1}^T \otimes \widehat{A}_N$ may be reduced by permutation and by sign scaling to a block diagonal matrix, where each block has at most N rows and takes one of the two following forms:

$$(6.8) \qquad J_k := \left.\begin{bmatrix} 1 & 1 & & \\ & 1 & \ddots & \\ & & \ddots & 1 \\ & & & 1 \end{bmatrix}\right\} k, \quad \widehat{J}_k := \left.\begin{bmatrix} 1 & 1 & & \\ & 1 & \ddots & \\ & & \ddots & 1 \\ & & & 1 & 1 \end{bmatrix}\right\} k.$$

The smallest singular value of $C(A_2, B_2)$ is given by the smallest singular value of the largest block of the permuted matrix, that is, either J_N or $\widehat{J}_N$, having dimensions $N \times N$ and $N \times (N+1)$, respectively. The smallest singular value of $C(A_2, B_2)$ is given by the smallest singular value of $J_N = I_N + \widehat{B}_N\widehat{A}_N^T$, where $\widehat{A}_N$ and $\widehat{B}_N$ are given

in (6.5). We bound the smallest singular value via the eigenvalues of $J_N J_N^T$. It turns out that this matrix is closely related to the colleague matrix that Good [11] proposed for Chebyshev polynomials of the second kind, and we may write the characteristic polynomial $\chi(\lambda) = \det\left(\lambda I + J_N J_N^T\right)$ as

$$(6.9) \qquad \det \begin{bmatrix} \lambda - 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \lambda - 2 & -1 \\ & & -1 & \lambda - 1 \end{bmatrix} = U_N(\lambda/2 - 1) + U_{N-1}(\lambda/2 - 1)\,,$$

where $U_N(\lambda)$ is the degree N Chebyshev polynomial of the second kind. This relation can be obtained by Laplace expansion along the last column of $\lambda I + J_N J_N^T$, combined with the recurrence relations for Chebyshev polynomials of the first and the second kind. Alternatively, with the change of variable $\mu = \lambda/2 - 1$, the matrix $\lambda I + J_N J_N^T$ becomes a colleague matrix for polynomial expansions in terms of Chebyshev polynomials of the second kind [11, p. 63].

What remains to be shown is that the smallest root of the characteristic polynomial $\chi(\lambda) = U_N(\lambda/2 - 1) + U_{N-1}(\lambda/2 - 1)$ lies between the two smallest roots of $T_{N+1}(\lambda/2 - 1)$ and $T_N(\lambda/2 - 1)$. With the change of variable $\mu = \lambda/2 - 1$ and combining a number of recurrence relations for Chebyshev polynomials (see, again, Good [11]), we rewrite $\chi(\lambda)$ as

$$(6.10) \qquad\qquad U_N(\mu) + U_{N-1}(\mu) = \frac{T_{N+1}(\mu) - T_N(\mu)}{\mu - 1}\,.$$

The roots of $T_N(\mu)$ interlace the roots of $T_{N+1}(\mu)$, and the polynomials have opposite signs at $\mu = 1$. Thus, the smallest root of $U_N(\mu) + U_{N-1}(\mu)$ must lie between the two smallest roots of $T_{N+1}(\mu)$ and $T_N(\mu)$. Expressed in terms of λ, these are

$$(6.11) \quad \lambda_{N+1} = 2\left(1 + \cos\left(\frac{(2N+1)\pi}{2(N+1)}\right)\right) \;\; \text{and} \;\; \lambda_N = 2\left(1 + \cos\left(\frac{(2N-1)\pi}{2N}\right)\right)\,.$$

Thus, the smallest singular values are the square roots of these values, and we can obtain the bounds (6.7) for the norm $\|C(A_2, B_2)^+\|_2$ by taking the reciprocal; this completes the proof. $\square$

COROLLARY 6.2. *Asymptotically, the norm $\|C(A_2, B_2)^+\|_2$ grows according to*

$$(6.12) \qquad\qquad \left\|C(A_2, B_2)^+\right\|_2 = \frac{2(N+1)}{\pi} + O\left(\frac{1}{N}\right)\,.$$

Proof. Simply take a series expansion of the upper bound in (6.7). $\square$

Remark 4. By applying the result of Jiang [14, Thm. 1], we can obtain upper and lower bounds for the singular values of J_N, somewhat similar to the bounds presented here. However, for our specific case, we have found that the bounds (6.7) are much tighter, with the asymptotic difference between the upper and lower bounds being a very modest $2/\pi$.

6.2. Chebyshev. The construction of the matrix $C(A_2, B_2)$ for the Chebyshev basis is slightly more complicated than for the monomial basis. Consider the following expansion:

$$(6.13) \qquad\qquad (\lambda B_2 - A_2)\Phi(\lambda) = \Theta T_{N+1}(\lambda) + \sum_{i=0}^{N} \Theta_i T_i(\lambda)\,,$$

where the $T_i(\lambda)$'s are Chebyshev polynomials of the first kind. We may write this product in terms of the coefficients of $\Phi(\lambda)$:

$$
(6.14) \qquad
\begin{bmatrix}
\Theta \\
\Theta_N \\
\vdots \\
\Theta_1 \\
\Theta_0
\end{bmatrix}
=
\begin{bmatrix}
\frac{1}{2}B_2 & & & & \\
-A_2 & \ddots & & & \\
\frac{1}{2}B_2 & \ddots & \frac{1}{2}B_2 & & \\
& \ddots & -A_2 & B_2 & \\
& & \frac{1}{2}B_2 & -A_2 &
\end{bmatrix}
\begin{bmatrix}
\Phi_N \\
\Phi_{N-1} \\
\vdots \\
\Phi_1 \\
\Phi_0
\end{bmatrix}.
$$

For the linearization given in (3.2), the matrix $C(A_2, B_2)$ also has the structure $A_2 = \widehat{A}_N \otimes I_s$ and $B_2 = \widehat{B}_N \otimes I_s$, where $\widehat{A}_N$ and $\widehat{B}_N$ are $N \times (N+1)$ matrices given by

$$
(6.15) \qquad
\widehat{A}_N =
\begin{bmatrix}
\frac{1}{2} & 0 & \frac{1}{2} & & \\
& \ddots & \ddots & \ddots & \\
& & \frac{1}{2} & 0 & \frac{1}{2} \\
& & & 1 & 0
\end{bmatrix},
\quad
\widehat{B}_N =
\begin{bmatrix} 0 & I_N \end{bmatrix}.
$$

Furthermore, one may write $C(A_2, B_2)$ in the form (6.1) with the above matrices, and where $\widehat{A}_{N+1}$ and $\widehat{B}_{N+1}$ have exactly the same structure as $\widehat{A}_N$ and $\widehat{B}_N$ but are of dimension $(N+1) \times (N+2)$.

We first symmetrize the linearization to simplify the analysis of the bounds for the singular values. Consider the following diagonal equivalence applied to the smaller matrix in (6.1):

$$
(6.16) \qquad (D_{N+2}^{-1} \otimes D_{N+1}^{-1})(\widehat{A}_{N+1}^T \otimes \widehat{B}_N - \widehat{B}_{N+1}^T \otimes \widehat{A}_N)(D_{N+1} \otimes D_N),
$$

where $D_k = \mathrm{diag}\{I_k, \sqrt{2}\}$. This modifies only the last row and column of $\widehat{A}_N$ and $\widehat{A}_{N+1}$, making them both equal to $1/\sqrt{2}$; furthermore, $\widehat{B}_N$ is unchanged by this operation. Thus, the symmetrized version of the matrix $C(A_2, B_2)$ is equal to

$$
(6.17) \qquad \widetilde{C}(A_2, B_2) = (\widetilde{A}_{N+1}^T \otimes \widehat{B}_N - \widehat{B}_{N+1}^T \otimes \widetilde{A}_N) \otimes I_s,
$$

where

$$
(6.18) \qquad
\widetilde{A}_N =
\begin{bmatrix}
\frac{1}{2} & 0 & \frac{1}{2} & & \\
& \ddots & \ddots & \ddots & \\
& & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} \\
& & & \frac{1}{\sqrt{2}} & 0
\end{bmatrix}.
$$

Note that we could also arrive at this same formulation simply by choosing a different normalization for the Chebyshev polynomials. In other words, if we make the normalization $T_0(\lambda) \mapsto \frac{1}{\sqrt{2}} T_0(\lambda)$ (see, for example, Trefethen [26, Chap. 17]), we obtain both the symmetrized version of the linearization and the matrix (6.17).

Compared with the monomial basis, the Chebyshev case is not as straightforward, and we are only able to obtain a lower bound for the norm $\|\widetilde{C}(A_2, B_2)^+\|_2$, as we will show in the following lemma.

LEMMA 6.3. *For the Chebyshev basis linearization* (3.2), *the norm of the pesudoinverse of* $C(A_2, B_2)$ *satisfies the lower bound*

$$
(6.19) \qquad \frac{1}{2\sqrt{2}\left(\cos\left(\frac{\pi}{2(N+1)}\right) - \cos\left(\frac{3\pi}{2(N+1)}\right)\right)} \leq \|C(A_2, B_2)^+\|_2.
$$

Proof. We use the symmetrized version of the matrix (6.17) to obtain our lower bound, since we make use of the symmetry of certain square submatrices. We will show that $\sigma_{\min}(\widehat{C}(A_2, B_2)) = \sigma_{N^2+2N}(\widetilde{A}_{N+1}^T \otimes \widehat{B}_N - \widehat{B}_{N+1}^T \otimes \widetilde{A}_N) \leq f(N)$ for some function f. To do this, we begin by constructing a vector $x \otimes y$ such that

$$(6.20) \quad \|(x^T \otimes y^T)(\widetilde{A}_{N+1}^T \otimes \widehat{B}_N - \widehat{B}_{N+1}^T \otimes \widetilde{A}_N)\|_2 = f(N)\|x^T \otimes y^T\|_2 = f(N)\|x\|_2\|y\|_2.$$

We construct x as $x^T = [\ 0 \quad \widehat{x}^T\ ]$, such that $\widehat{B}_{N+1}x = \widehat{x}$, $\|\widehat{x}\|_2 = \|x\|_2 = 1$, and $\widehat{A}_{N+1}x = \widetilde{A}\widehat{x} = \widetilde{\lambda}\widehat{x}$, where $\widetilde{\lambda}$ is the smallest (or largest) eigenvalue of the $(N+1) \times (N+1)$ matrix

$$(6.21) \qquad \widetilde{A} = \begin{bmatrix} 0 & \frac{1}{2} & & & \\ \frac{1}{2} & \ddots & \ddots & & \\ & \ddots & \ddots & \frac{1}{2} & \\ & & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} \\ & & & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}.$$

It thus follows that

$$(6.22) \quad (x^T \otimes y^T)(\widetilde{A}_{N+1}^T \otimes \widehat{B}_N - \widehat{B}_{N+1}^T \otimes \widetilde{A}_N)$$
$$= \widetilde{\lambda}\widehat{x}^T \otimes y^T \widehat{B}_N - \widehat{x}^T \otimes y^T \widetilde{A}_N = \widehat{x}^T \otimes y^T(\widetilde{\lambda}\widehat{B}_N - \widetilde{A}_N).$$

Let us now choose a vector y such that $\|y\|_2 = 1$, and which minimizes

$$(6.23) \qquad \|y^T(\widetilde{\lambda}\widehat{B}_N - \widetilde{A}_N)\|_2 = \sigma_{\min}(\widetilde{\lambda}\widehat{B}_N - \widetilde{A}_N) =: f(N).$$

We then have that $\|x \otimes y\|_2 = 1$, $\|\widetilde{C}(A_2, B_2)^T((x \otimes y) \otimes I_s)\|_2 = f(N)$, and hence $\sigma_{\min}(\widetilde{C}(A_2, B_2)) \leq f(N)$.

We now use a result by Thompson [24] to bound the singular values by embedding the matrix of interest inside a larger matrix. That is, we may embed $\widetilde{\lambda}\widehat{B}_N - \widetilde{A}_N$ in the larger matrix

$$(6.24) \qquad \widetilde{\lambda}I_{N+1} - \widetilde{A} = \begin{bmatrix} \widetilde{\lambda}e_1^T - e_2^T/2 \\ \widetilde{\lambda}\widehat{B}_N - \widetilde{A}_N \end{bmatrix},$$

where e_1 and e_2 are the first and second unit vectors of length $N + 1$, respectively. Since $\widetilde{\lambda}$ is already an eigenvalue of $\widetilde{A}$, the object of interest is the gap between the two smallest singular values of $\widetilde{A}$. Recall that $\widetilde{A}$ is symmetric, and we can thus compute the singular values via the eigenvalues of $\widetilde{A}$. We can immediately identify that the characteristic polynomial of $\widetilde{A}$ is the Chebyshev polynomial $T_{N+1}(\lambda)$. This can be derived by applying diagonal similarity transformations to the colleague matrix introduced by Good [11]. The eigenvalues of $\widetilde{A}$ are the zeros of the Chebyshev polynomial $T_{N+1}(\lambda)$, which are

$$(6.25) \qquad \lambda_k = \cos\left(\frac{2k-1}{2(N+1)}\pi\right), \quad 1 \leq k \leq N+1.$$

Therefore, Thompson's inequality [24] implies that

$$(6.26) \qquad 0 \leq \sigma_{\min}(\widetilde{\lambda}\widehat{B}_N - \widetilde{A}_N) = f(N) \leq \cos\left(\frac{\pi}{2(N+1)}\right) - \cos\left(\frac{3\pi}{2(N+1)}\right).$$

Finally, we deduce that the norm of the pseudoinverse $C(A_2, B_2)^+$ is bounded from below by

$$(6.27) \qquad \frac{1}{2\sqrt{2}\left(\cos\left(\frac{\pi}{2(N+1)}\right) - \cos\left(\frac{3\pi}{2(N+1)}\right)\right)} \leq \|C(A_2, B_2)^+\|_2 . \qquad \Box$$

Remark 5. For large N, the lower bound (6.27) behaves as

$$(6.28) \qquad \frac{(N+1)^2}{\pi^2\sqrt{2}} + O\left(\frac{1}{N}\right) + O(1) ,$$

asymptotically.

6.3. Lagrange. Unlike the Monomial and Chebyshev bases, the Lagrange basis is not degree graded, and thus there is some additional choice in representing polynomials of degree $N + 1$. We could add an additional node, but this would modify all of the Lagrange basis elements. Instead, we make use of the node polynomial $\ell(\lambda)$ as our extra basis element (since it is already linearly independent), and thus we consider the expansion

$$(6.29) \qquad (\lambda B_2 - A_2)\Phi(\lambda) = \Theta\ell(\lambda) + \sum_{i=0}^{N} \Theta_i\ell_i(\lambda) .$$

Evaluating the expansion at each of the nodes in turn, we immediately recover the coefficients $\Theta_i = (\sigma_i B_2 - A_2)\Phi_i$. The coefficient Θ is obtained by substituting the computed Θ_i's into (6.29) and then using the relation between the barycentric weights, the Lagrange polynomials, and the node polynomial given in (3.4). The resulting linear system is given by

$$(6.30) \qquad \begin{bmatrix} \Theta \\ \Theta_N \\ \vdots \\ \Theta_0 \end{bmatrix} = \begin{bmatrix} \beta_N B_2 & \cdots & \beta_0 B_2 \\ \sigma_N B_2 - A_2 & & \\ & \ddots & \\ & & \sigma_0 B_2 - A_2 \end{bmatrix} \begin{bmatrix} \Phi_N \\ \Phi_{N-1} \\ \vdots \\ \Phi_0 \end{bmatrix} ,$$

which can also be written in the form (6.1), where $\widehat{A}_N$ and $\widehat{B}_N$ can be identified from the linearization (3.5) as

$$(6.31)$$
$$\widehat{A}_N = \begin{bmatrix} \sigma_N & -\sigma_{N-1}\theta_{N-1} & \\ & \ddots & \ddots \\ & & \sigma_1 & -\sigma_0\theta_0 \end{bmatrix} , \quad \widehat{B}_N = \begin{bmatrix} 1 & -\theta_{N-1} & \\ & \ddots & \ddots \\ & & 1 & -\theta_0 \end{bmatrix} ,$$

and where $\widehat{A}_{N+1}$ and $\widehat{B}_{N+1}$ can be identified from (6.30) as

$$(6.32) \qquad \widehat{A}_{N+1} = \begin{bmatrix} \beta_N & \sigma_N & \\ \vdots & & \ddots \\ \beta_0 & & \sigma_0 \end{bmatrix} , \quad \widehat{B}_{N+1} = \begin{bmatrix} 0 & I_{N+1} \end{bmatrix} .$$

One may further identify that $\widehat{A}_N = \widehat{B}_N\Sigma$, where $\Sigma = \text{diag}\{ \sigma_N \ \cdots \ \sigma_0 \}$; hence, we may write $C(A_2, B_2)$ as

$$(6.33) \qquad C(A_2, B_2) = \left(\begin{bmatrix} \beta^T \\ \Sigma \end{bmatrix} \otimes \widehat{B}_N - \begin{bmatrix} 0 \\ I_{N+1} \end{bmatrix} \otimes B_N\Sigma \right) \otimes I_s ,$$

where $\beta^T = [\ \beta_N \ \cdots \ \beta_0 \ ]$ is the vector of barycentric weights.

A lower bound for the norm $\|C(A_2, B_2)^+\|_2$ can be obtained by following a similar strategy as for the Chebyshev case: we construct a vector to show that $\sigma_{\min}(C(A_2, B_2)) \leq f(N)$. It turns out that there is a particularly simple choice for the structured vector, allowing us to obtain an approximate upper bound for the smallest singular value, and thus we obtain a lower bound for the norm of the pseudoinverse.

LEMMA 6.4. *For the Lagrange linearization* (3.5)*, we obtain the following lower bound:*

$$(6.34) \qquad \frac{1}{|\theta_{N-k}(\sigma_{N-k+1} - \sigma_{N-k})|} \leq \|C(A_2, B_2)^+\|_2$$

for $1 \leq k \leq N$.

Proof. Since $C(A_2, B_2)$ given in (6.33) is a Kronecker product, we only need to compute a bound for the singular values of the smaller matrix in the product. Thus, we construct a vector $x \otimes y$, where $x^T = [\ 0 \quad e_k^T\ ] \in \mathbb{R}^{N+2}$ and $y^T = \widehat{e}_k^T \in \mathbb{R}^N$ are unit vectors, such that we have the following string of equalities:

$$(6.35)\ (x^T \otimes y^T) \left(\begin{bmatrix} \beta^T \\ \Sigma \end{bmatrix} \otimes B_N - \begin{bmatrix} 0 \\ I_{N+1} \end{bmatrix} \otimes \widehat{B}_N \Sigma \right)$$

$$= \left(\sigma_{N-k+1} e_k^T \otimes \widehat{e}_k^T \widehat{B}_N - e_k^T \otimes \widehat{e}_k^T B_N \Sigma \right) = e_k^T \otimes \widehat{e}_k^T \widehat{B}_N (\sigma_{N-k+1} I - \Sigma)$$

$$= e_k^T \otimes [(e_k^T - \theta_{N-k} e_{k+1}^T)(\sigma_{N-k+1} I - \Sigma)]$$

$$= -\theta_{N-k}(\sigma_{N-k+1} - \sigma_{N-k})(e_k \otimes e_{k+1})^T .$$

We thus deduce that $\|((x^T \otimes y^T) \otimes I_s) C(A_2, B_2)\|_2 = |\theta_{N-k}(\sigma_{N-k+1} - \sigma_{N-k})|$ and therefore that $\sigma_{\min}(C(A_2, B_2)) \leq |\theta_{N-k}(\sigma_{N-k+1} - \sigma_{N-k})|$, completing the proof. $\square$

Let us consider the ramifications of this result for two specific choices for the interpolation nodes, namely, the roots of unity and Chebyshev nodes of the second kind.

Remark 6. For the Lagrange basis linearization (3.5), we obtain the following bounds for specific choices of nodes:

- For the roots of unity $\sigma_k = \exp\left(\frac{2\pi i k}{N+1}\right)$, $0 \leq k \leq N$, we have

$$(6.36) \qquad \frac{1}{|1 - \exp\left(\frac{2\pi i N}{N+1}\right)|} \leq \|C(A_2, B_2)^+\|_2 ,$$

and asymptotically this bound grows like $O(N)$.

- For Chebyshev nodes of the second kind $\sigma_k = \cos\left(\frac{k\pi}{N}\right)$, $0 \leq k \leq N$, we have

$$(6.37) \qquad \frac{2}{|1 - \cos\left(\frac{\pi}{N}\right)|} \leq \|C(A_2, B_2)^+\|_2 ,$$

and asymptotically this bound grows like $O(N^2)$.

Proof. For the roots of unity, all of the $\theta_i = \beta_{i+1}/\beta_i$ have unit magnitude, and thus the lower bound is driven by the distance between the last two nodes; a series expansion shows that this is of order $O(N)$. For the Chebyshev nodes of the second kind, the ratios θ_i all have unit magnitude except for the first and the last, which have magnitude 2 and 1/2, respectively. Again, the lower bound is driven by the distance between the last two nodes, and this behaves like $O(N^2)$ asymptotically. $\square$

TABLE 1
Upper bounds for the norm $\|\widehat{A}_{N+1}\|_2$.

Basis	$\|\widehat{A}_{N+1}\|_2$
Monomial	1
Chebyshev	$\leq \sqrt{4/3}$
Unit points	$\leq \sqrt{1 + 1/(N+1)}$
Chebyshev points	$\leq \sqrt{N+1}$

Remark 7. In sections 6.2 and 6.3, we have only been able to obtain lower bounds for the norm $\|C(A_2, B_2)^+\|_2$. These lower bounds are, however, a good approximation to the actual values and thus provide useful insight into the growth of this specific factor. We remark further that the singular values of $C(A_2, B_2)$ are independent of the specific polynomial coefficients.

7. Numerical experiment. One of the primary objectives of this work is to provide useful and easily computable upper bounds for the backward error of the solution of the polynomial eigenvalue problem. This enables us to indicate when we can compute the solution of the polynomial eigenvalue problem in a backward-stable way. Thus far, we have introduced the upper bound (5.9) that depends on the choice of basis used to construct the linearization. The upper bound depends primarily on the norm of the polynomial coefficients $\|A_1\|_2 = \|[\ P_N\ \cdots\ P_0\ ]\|_2$, the norm of the pseudoinverse $\|C(A_2, B_2)^+\|_2$, and the backward errors from the QZ algorithm. For the linearizations that we introduced in section 3, we developed bounds for the growth of $\|C(A_2, B_2)^+\|_2$, and now we turn to the remaining terms in the upper bound (5.9). The two terms that remain involve the matrices $C(A_2, B_2)$ based on the specific choice of basis. We were able to write $C(A_2, B_2)$ as the sum of two Kronecker products, and in a similar way, we can write the two remaining matrices occurring in (5.9) as

$$(7.1) \qquad C(\Delta A_i, \Delta B_i) = \widehat{A}_{N+1}^T \otimes \Delta B_i - \widehat{B}_{N+1}^T \otimes \Delta A_i, \quad i = 1, 2.$$

We can then form the larger upper bound for the block vector of coefficients of $\Delta P(\lambda)$:

$$(7.2)$$
$$\left\|\overrightarrow{\Delta P}\right\|_2 \leq \|(\Delta A, \Delta B)\|_F \|\Psi\|_2 \left(\|\widehat{A}_{N+1}^T\|_2 + \|\widehat{B}_{N+1}^T\|_2\right) \left(1 + \|A_1\|_2 \|C(A_2, B_2)^+\|_2\right).$$

This larger upper bound has the advantage that we can compute it directly from the linearization and the bound for the backward error from the QZ algorithm (that is, using the bound on the perturbations $\|(\Delta A, \Delta B)\|_F$). For all of the linearizations defined in section 3, inspecting the coefficients of $\Phi(\lambda)$ reveals that $\Psi = \text{vec}(I_{N+1}) \otimes I_s$ and thus $\|\Psi\|_2 = \sqrt{N+1}$. Furthermore, we have $\|\widehat{B}_{N+1}\|_2 = 1$, and upper bounds for the values of $\|\widehat{A}_{N+1}\|_2$ are given in Table 1. For each of the bases considered, the bounds follow from the inequality $\|\widehat{A}_{N+1}\|_2 \leq \sqrt{\|\widehat{A}_{N+1}\|_1 \|\widehat{A}_{N+1}\|_\infty}$. It suffices to note that for points on the unit circle we have $\beta_k = \sigma_k/(N+1)$, and for Chebyshev points of the second kind we have $\beta_k = (-1)^k$ for $1 \leq k \leq N-1$ and $\beta_0 = 1/2$, $\beta_N = (-1)^N/2$.

At this point we now have an upper bound for the backward error that depends on the degree of the polynomial and on the norm of the polynomial coefficients $\|A_1\|_2$. Thus, we wish to investigate polynomials where we will be able see how close our theoretical bounds are to the computed backward errors. Hence, we investigate very well-behaved polynomials, which will exemplify the degree-dependent behavior for
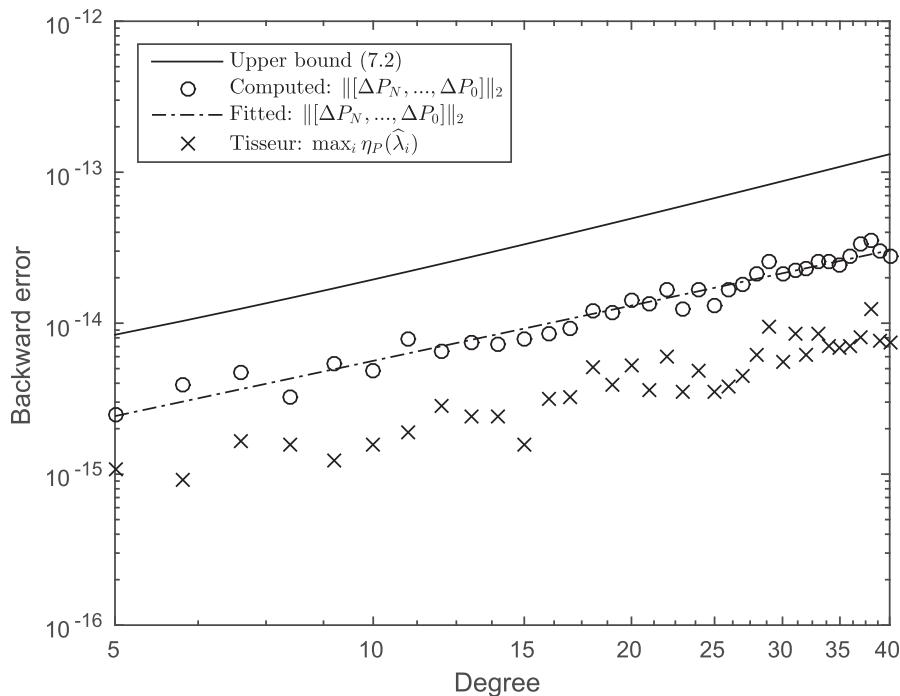
Fig. 1. *Monomial basis linearization.*

each of the linearizations discussed in this manuscript. By "well behaved," we mean that all of the polynomial coefficient norms $\|P_i\|_2$ are all equal. According to Tisseur [25, Thm. 7] (for quadratic eigenvalue problems), and Higham, Li, and Tisseur [13] (for higher degree polynomials), the condition that all $\|P_i\|_2 \approx 1$ implies that the backward errors of the polynomial eigenvalue problem are approximately equal to the backward errors of the linearization.

We found that by drawing coefficients from the circular unitary ensemble, and scaling the coefficients so that $\|[\; P_N \quad \cdots \quad P_0 \;]\|_2 = 1$, the resulting polynomials are well behaved. The polynomials have eigenvalues that are close to the underlying measure associated with the polynomial basis used. Thus, for the monomial basis the eigenvalues cluster near the unit circle, for the Chebyshev basis the eigenvalues cluster near the unit interval, and for the Lagrange basis, the eigenvalues cluster near the interpolation nodes.

Our experiment was performed on polynomials having dimension $s = 2$ and increasing degree up to 40. The backward errors in the polynomial coefficients were computed using (5.8) where the perturbations ΔA and ΔB were reconstructed from the computed eigenvalues and eigenvectors produced by the function qz in MATLAB. For all linearizations the coefficient $\Delta P_{N+1} = 0$, and thus we recover the s infinite eigenvalues exactly. We compared the backward errors with the bound (7.2), and show this in Figures 1–4. For the monomial basis, we also include a comparison with the bound given by Higham, Li, and Tisseur [13], that is, the expression given in (4.4), where we take the maximum over all of the computed eigenvalues. Figure 1 shows that the global backward error bound that we present here exceeds the local one only by an order of magnitude. Since our bound is a global bound, rather than just for specific eigenpairs, we find this overestimation to be quite satisfactory. Table 2 shows
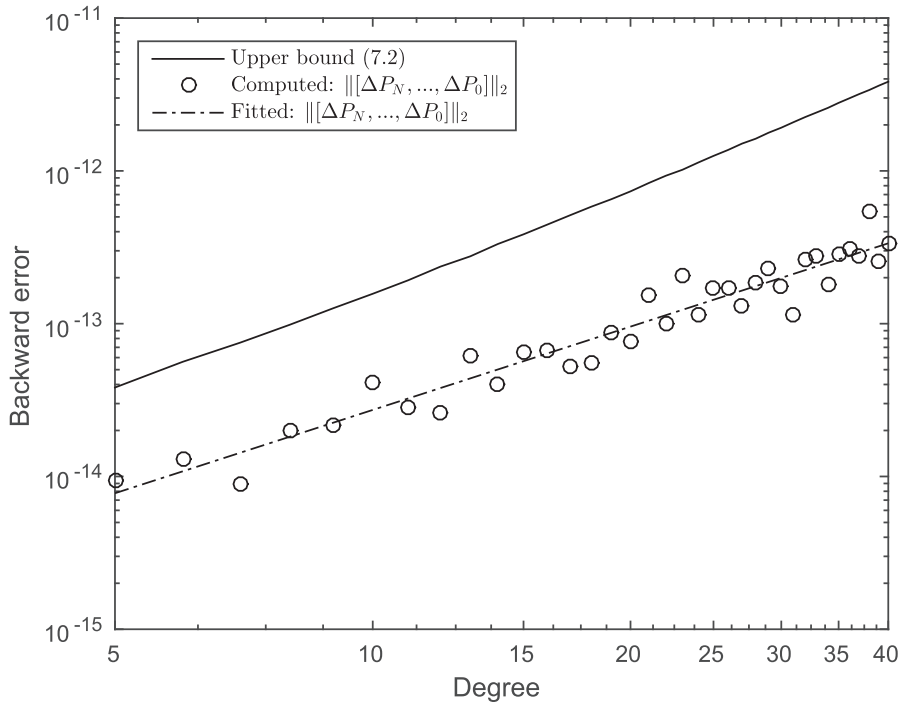
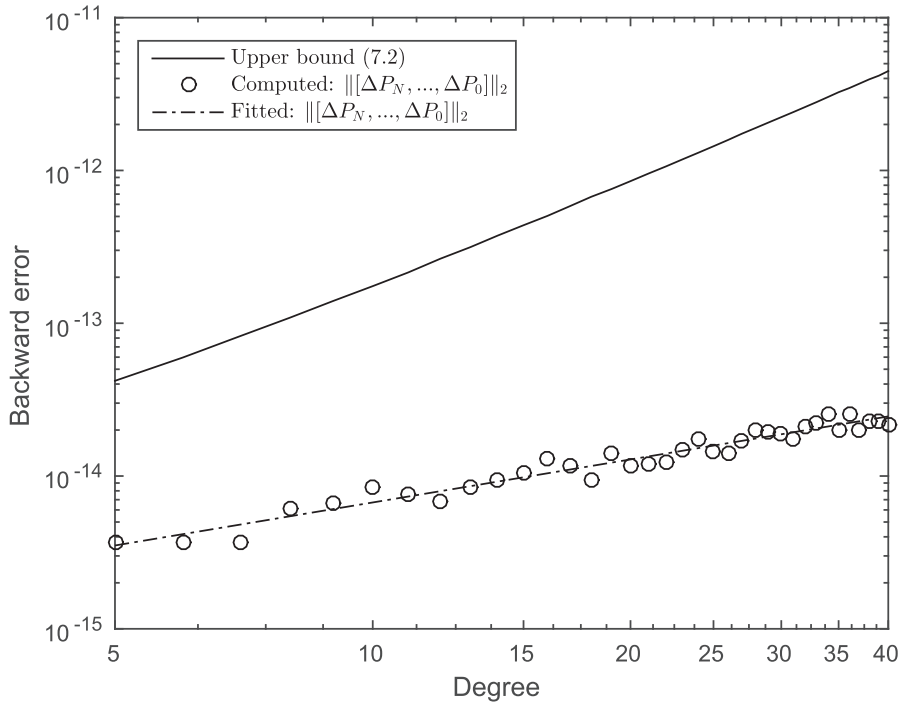FIG. 2. *Chebyshev basis linearization.*

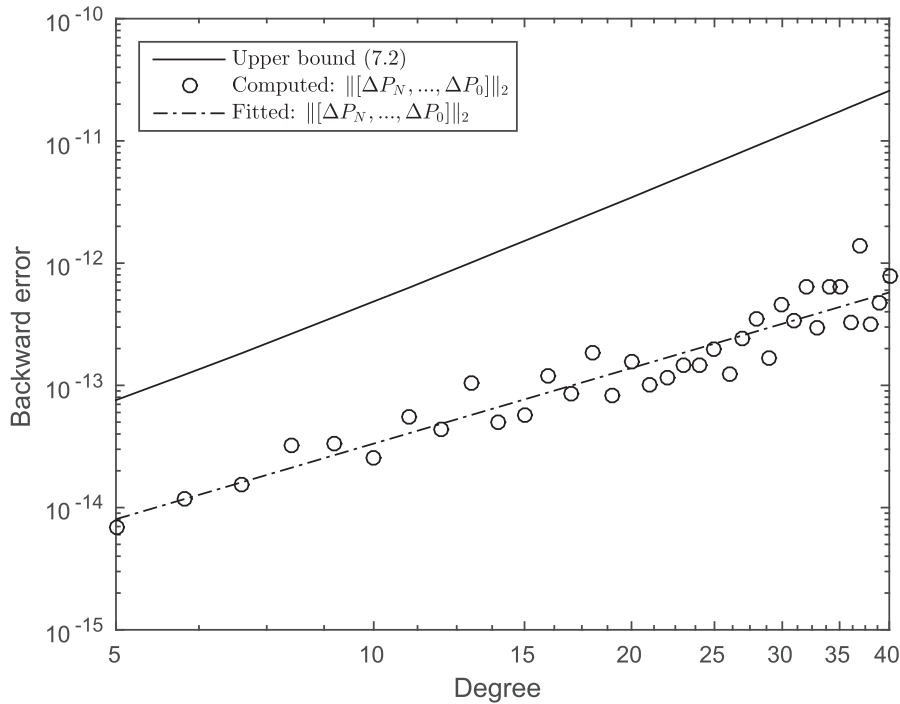FIG. 3. *Lagrange basis linearization: Unit circle points.*

FIG. 4. *Lagrange basis linearization: Chebyshev nodes.*

TABLE 2
Comparison of growth factors.

Basis	Fitted growth	Bound growth
Monomial	1.2	1.3
Chebyshev	1.8	2.2
Unit points	0.9	2.2
Chebyshev points	2.1	2.8

the exponents of the growth of the upper bound and backward error with respect to the degree, that is, we tabulate the exponent α of the model βN^α fitted in the least squares sense to the computed backward errors and upper bound.

For the Monomial and Chebyshev linearizations, we see that the upper bound is exceeding the computed backward errors only by around one order of magnitude. The growth rate is also captured quite accurately. For the Lagrange basis, we interpolate at two sets of nodes: the roots of unity and Chebyshev points of the first kind. For the roots of unity, we see something quite unusual, the upper bound overestimates the growth in the backward error by more than a factor of N. The overestimation of the growth is less pronounced for the Chebyshev points.

8. Discussion and conclusion. In this manuscript, we have introduced a procedure for constructing strong linearizations of order $(N+1)s$ of polynomial matrices. This construction allows for a uniform backward error analysis over a wide range of linearizations expressed in different polynomial bases. We have shown that the growth in the backward errors is dependent on only a small number of factors: the norm of the pseudoinverse of the matrix $C(A_2, B_2)$, the norm of the block vector of polynomial

coefficients, the norm of the coefficients of the one-sided factorization, and the norms of $\widehat{A}_{N+1}$ and $\widehat{B}_{N+1}$ that are related to the polynomial basis. For specific bases and linearizations we have shown that the upper bound is a reasonable approximation of the computed backward error for well-behaved problems, and for the classical monomial and Chebyshev linearizations it accurately captures the growth in the backward error with increasing degree.

REFERENCES

[1] A. AMIRASLANI, R. M. CORLESS, AND P. LANCASTER, *Linearization of matrix polynomials expressed in polynomial bases*, IMA J. Numer. Anal., 29 (2009), pp. 141–157.

[2] E. ANDERSON, Z. BAI, C. BISCHOF, S. BLACKFORD, J. DEMMEL, J. DONGARRA, J. DU CROZ, A. GREENBAUM, S. HAMMARLING, A. MCKENNEY, AND D. SORENSEN, *LAPACK Users' Guide*, 3rd ed., SIAM, Philadelphia, 1999.

[3] J.-P. BERRUT AND L. N. TREFETHEN, *Barycentric Lagrange interpolation*, SIAM Rev., 46 (2004), pp. 501–517.

[4] R. M. CORLESS, *Generalized companion matrices in the Lagrange basis*, in Proceedings of EACA, L. Gonzalez-Vega and T. Recio, eds., 2004, pp. 317–322.

[5] F. DE TERÁN, F. M. DOPICO, AND J. PÉREZ, *Backward stability of polynomial root-finding using Fiedler companion matrices*, IMA J. Numer. Anal., 2015.

[6] F. DE TERÁN, F. DOPICO, AND P. VAN DOOREN, *Constructing strong ℓ-ifications from dual minimal bases*, Linear Algebra Appl., to appear.

[7] A. EDELMAN AND H. MURAKAMI, *Polynomial roots from companion matrix eigenvalues*, Math. Comp., 64 (1995), pp. 763–776.

[8] M. FIEDLER, *Expressing a polynomial as the characteristic polynomial of a symmetric matrix*, Linear Algebra Appl., 141 (1990), pp. 265–270.

[9] G. D. FORNEY, JR., *Minimal bases of rational vector spaces, with applications to multivariable linear systems*, SIAM J. Control, 13 (1975), pp. 493–520.

[10] I. C. 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.

[11] I. J. GOOD, *The colleague matrix, a Chebyshev analogue of the companion matrix*, Q. J. Math., 12 (1961), pp. 61–68.

[12] L. GRAMMONT, N. J. HIGHAM, AND F. TISSEUR, *A framework for analyzing nonlinear eigenproblems and parametrized linear systems*, Linear Algebra Appl., 435 (2011), pp. 623–640.

[13] N. J. HIGHAM, R.-C. LI, AND F. TISSEUR, *Backward error of polynomial eigenproblems solved by linearization*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1218–1241.

[14] E. JIANG, *Bounds for the smallest singular value of a Jordan block with an application to eigenvalue perturbation*, Linear Algebra Appl., 197 (1994), pp. 691–707.

[15] T. KAILATH, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[16] 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, Manchester, UK, 2006.

[17] P. LANCASTER, *Linearization of regular matrix polynomials*, Electron. J. Linear Algebra, 17 (2008), pp. 21–27.

[18] P. W. LAWRENCE AND R. M. CORLESS, *Stability of rootfinding for barycentric Lagrange interpolants*, Numer. Algorithms, 65 (2014), pp. 447–464.

[19] P. W. LAWRENCE AND R. M. CORLESS, *Backward error of polynomial eigenvalue problems solved by linearization of Lagrange interpolants*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1425–1442.

[20] F. MALEK AND R. VAILLANCOURT, *Polynomial zerofinding iterative matrix algorithms*, Comput. Math. Appl., 29 (1995), pp. 1–13.

[21] Y. NAKATSUKASA AND V. NOFERINI, *On the stability of computing polynomial roots via confederate linearizations*, Math. Comp., (2015), `doi:10.1090:mcom3039`.

[22] A. C. R. NEWBERY, *A family of test matrices*, Commun. ACM, 7 (1964), pp. 724–724.

[23] V. NOFERINI AND J. PÉREZ ALVARO, *Chebyshev Rootfinding via Computing Eigenvalues of Colleague Matrices: When Is It Stable?*, MIMS EPrint 2015.25, Manchester Institute for Mathematical Sciences, The University of Manchester, Manchester, UK, 2015.

[24] R. C. THOMPSON, *Interlacing inequalities for invariant factors*, Linear Algebra Appl., 24 (1979), pp. 1–31.

[25] F. TISSEUR, *Backward error and condition of polynomial eigenvalue problems*, Linear Algebra Appl., 309 (2000), pp. 339–361.

[26] L. N. TREFETHEN, *Approximation Theory and Approximation Practice*, SIAM, Philadelphia, 2013.

[27] R. VAN BEEUMEN, W. MICHIELS, AND K. MEERBERGEN, *Linearization of Lagrange and Hermite interpolating matrix polynomials*, IMA J. Numer. Anal., 35 (2014), pp. 909–930.

[28] P. VAN DOOREN AND P. DEWILDE, *The eigenstructure of an arbitrary polynomial matrix: Computational aspects*, Linear Algebra Appl., 50 (1983), pp. 545–579.