Vector spaces of linearizations for matrix polynomials: a bivariate polynomial approach

Nakatsukasa, Yuji and Noferini, Vanni and Townsend, Alex

2012

MIMS EPrint: **2012.118**

ISSN 1749-9097

VECTOR SPACES OF LINEARIZATIONS FOR MATRIX POLYNOMIALS: A BIVARIATE POLYNOMIAL APPROACH

YUJI NAKATSUKASA*, VANNI NOFERINI†, AND ALEX TOWNSEND‡

In memory of Leiba Rodman

Abstract. We revisit the landmark paper [D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 971–1004] and, by viewing matrices as coefficients for bivariate polynomials, we provide concise proofs for key properties of linearizations for matrix polynomials. We also show that every pencil in the double ansatz space is intrinsically connected to a Bézout matrix, which we use to prove the eigenvalue exclusion theorem. In addition our exposition allows for any polynomial basis and for any field. The new viewpoint also leads to new results. We generalize the double ansatz space by exploiting its algebraic interpration as a space of Bézout pencils to derive new linearizations with potential applications in the theory of structured matrix polynomials. Moreover, we analyze the conditioning of double ansatz space linearization in the important practical case of a Chebyshev basis.

Key words. matrix polynomials; bivariate polynomials; Bézoutian; double ansatz space; degree-graded polynomial basis; orthogonal polynomials; conditioning

AMS subject classifications. 65F15, 15A18, 15A22

1. Introduction.

The landmark paper by Mackey, Mackey, Mehl, and Mehrmann [20] introduced three important vector spaces of pencils for matrix polynomials: $\mathbb{L}_1(P)$, $\mathbb{L}_2(P)$, and $\mathbb{DL}(P)$. In [20] the spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ generalize the companion forms of the first and second kind, respectively, and the *double ansatz space* is the intersection, $\mathbb{DL}(P) = \mathbb{L}_1(P) \cap \mathbb{L}_2(P)$.

In this article we introduce new viewpoints for these vector spaces, which are important for solving polynomial eigenvalue problems. The classic approach is *linearization*, i.e., computing the eigenvalues of a matrix polynomial $P(\lambda)$ by solving a generalized linear eigenvalue problem. The vector spaces we study provide a family of candidate generalized eigenvalue problems for computing the eigenvalues of a matrix polynomial. We regard a block matrix as coefficients for a bivariate matrix polynomial (see section 3), and point out that every pencil in $\mathbb{DL}(P)$ is a (generalized) Bézout matrix [18] (see section 4). These novel viewpoints allow us to obtain remarkably elegant proofs for many properties of $\mathbb{DL}(P)$ and the eigenvalue exclusion theorem, which previously required rather tedious derivations. Furthermore, our exposition includes matrix polynomials expressed in any polynomial basis, such as the Chebyshev polynomial [8, 17]. We develop a generalization of the double ansatz space (see section 5) and also discuss extensions to generic algebraic fields, and conditioning analysis (see section 6).

Let us recall some basic definitions in the theory of matrix polynomials. Let $P(\lambda) = \sum_{i=0}^{k} A_i \phi_i(\lambda)$ be a matrix polynomial expressed in a certain polynomial basis $\{\phi_0, \ldots, \phi_k\}$, where $A_k \neq 0$, $A_i \in \mathbb{F}^{n\times n}$, and $\mathbb{F}$ is a field. Of particular interest is

*Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Tokyo 113-8656, Japan (nakatsukasa@mist.i.u-tokyo.ac.jp). Supported by EPSRC grant EP/I005293/1 and JPS grant No. 26870149.

†School of Mathematics, University of Manchester, Manchester, England, M13 9PL (vanni.noferini@manchester.ac.uk). Supported by European Research Council Advanced Grant MATFUN (267526).

‡Mathematics Department, Massachussets Institute of Technology, 77 Massachussets Ave, Cambridge, MA 02139 (ajt@mit.edu). .

the case of a degree-graded basis, i.e., $\{\phi_i\}$ is a polynomial basis where ϕ_j is of exact degree j. We assume throughout that $P(\lambda)$ is regular, i.e., $\det P(\lambda) \not\equiv 0$, which ensures the finite eigenvalues of $P(\lambda)$ are the roots of the scalar polynomial $\det(P(\lambda))$. We note that if the elements of A_i are in the field $\mathbb{F}$ then generally the finite eigenvalues exist in the algebraic closure of $\mathbb{F}$.

Given $X, Y \in \mathbb{F}^{nk\times nk}$ a matrix pencil $L(\lambda) = \lambda X + Y$ is a *linearization* for $P(\lambda)$ if there exist unimodular matrix polynomials $U(\lambda)$ and $V(\lambda)$, i.e., $\det U(\lambda), \det V(\lambda)$ are nonzero elements of $\mathbb{F}$, such that $L(\lambda) = U(\lambda)\operatorname{diag}(P(\lambda), I_{n(k-1)})V(\lambda)$ and hence, $L(\lambda)$ shares its finite eigenvalues and their partial multiplicities with $P(\lambda)$. If $P(\lambda)$, when expressed in a degree-graded basis, has a singular leading coefficient then it has an infinite eigenvalue and to preserve the partial multiplicities at infinity the matrix pencil $L(\lambda)$ needs to be a *strong linearization*, i.e., $L(\lambda)$ is a linearization for $P(\lambda)$ and $\lambda Y + X$ a linearization for $\lambda^k P(1/\lambda)$.

In the next section we recall the definitions of $\mathbb{L}_1(P)$, $\mathbb{L}_2(P)$, and $\mathbb{DL}(P)$ allowing for matrix polynomials expressed in any polynomial basis. In section 3 we consider the same space from a new viewpoint, based on bivariate matrix polynomials, and provide concise proofs for properties of $\mathbb{DL}(P)$. Section 4 shows that every pencil in $\mathbb{DL}(P)$ is a (generalized) Bézout matrix and gives an alternative proof for the eigenvalue exclusion theorem. In section 5 we generalize the double ansatz space to obtain a new family of linearizations, including new structured linearizations for structured matrix polynomials. Although these new linearizations are mainly of theoretical interest they show how the new viewpoint can be used to derive novel results. In section 6 we analyze the conditioning of the eigenvalues of $\mathbb{DL}(P)$ pencils, and in section 7 we describe a procedure to construct block symmetric pencils in $\mathbb{DL}(P)$ and Bézout matrices.

2. Vector spaces and polynomial bases. Given a matrix polynomial $P(\lambda)$ we can define a vector space, denoted by $\mathbb{L}_1(P)$, as [20, Def. 3.1]

$$\mathbb{L}_1(P) = \left\{ L(\lambda) = \lambda X + Y : X, Y \in \mathbb{F}^{nk\times nk},\ L(\lambda)\cdot(\Lambda(\lambda)\otimes I_n) = v \otimes P(\lambda), v \in \mathbb{F}^k \right\},$$

where $\Lambda(\lambda) = [\phi_{k-1}(\lambda), \phi_{k-2}(\lambda), \ldots, \phi_0(\lambda)]^T$ and $\otimes$ is the matrix Kronecker product. An *ansatz vector* $v \in \mathbb{F}^k$ generates a family of pencils in $\mathbb{L}_1(P)$, which are generically linearizations for $P(\lambda)$ [20, Thm. 4.7]. If $\{\phi_0, \ldots, \phi_k\}$ is an orthogonal basis, then the comrade form [26] belongs to $\mathbb{L}_1(P)$ with $v = [1, 0, \ldots, 0]^T$.

The action of $L(\lambda) = \lambda X + Y \in \mathbb{L}_1(P)$ on $(\Lambda(\lambda)\otimes I_n)$ can be characterized by the *column shift sum* operator, denoted by $\boxplus\!\!\rightarrow$ [20, Lemma 3.4],

$$L(\lambda)\cdot(\Lambda(\lambda)\otimes I_n) = v\otimes P(\lambda) \Longleftrightarrow X \boxplus\!\!\rightarrow Y = v \otimes [A_k, A_{k-1}, \ldots, A_0].$$

In the monomial basis $X \boxplus\!\!\rightarrow Y$ can be paraphrased as "insert a zero column on the right of X and a zero column on the left of Y then add them together", i.e.,

$$X \boxplus\!\!\rightarrow Y = \begin{bmatrix} X & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & Y \end{bmatrix},$$

where $\mathbf{0} \in \mathbb{F}^{nk\times n}$. More generally, given a polynomial basis we define the column shift sum operator as

$$X \boxplus\!\!\rightarrow Y = XM + \begin{bmatrix} \mathbf{0} & Y \end{bmatrix}, \tag{2.1}$$

where $M \in \mathbb{F}^{nk\times n(k+1)}$ and $\mathbf{0} \in \mathbb{F}^{nk\times n}$. The matrix M has a particularly nice form if the basis is degree-graded. Indeed, suppose the degree-graded basis $\{\phi_0, \ldots, \phi_k\}$

satisfies the recurrence relations

$$x\phi_{i-1} = \sum_{j=0}^{i} m_{k+1-i,k+1-j}\phi_j, \quad 1 \le i \le k.$$

Then the matrix M in (2.1) is given by

$$M = \begin{bmatrix} M_{11} & M_{12} & \dots & M_{1k} & M_{1,k+1} \\ \mathbf{0} & M_{22} & \ddots & \ddots & M_{2,k+1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & M_{kk} & M_{k,k+1} \end{bmatrix}, \tag{2.2}$$

where $M_{pq} = m_{pq}I_n$, $1 \le p \le q \le k+1$, $p \ne k+1$ and I_n is the $n \times n$ identity matrix. An orthogonal basis satisfies a three term recurrence and in this case the matrix M has only three nonzero block diagonals. For example, if $P(\lambda) \in \mathbb{R}[\lambda]^{n\times n}$ is expressed in the Chebyshev basis[1] $\{T_0(x), \dots, T_k(x)\}$, where $T_j(x) = \cos\left(j \cos^{-1} x\right)$ for $x \in [-1, 1]$, we have

$$M = \begin{bmatrix} \frac{1}{2}I_n & 0 & \frac{1}{2}I_n & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{1}{2}I_n & 0 & \frac{1}{2}I_n \\ & & & I_n & 0 \end{bmatrix} \in \mathbb{R}^{nk\times n(k+1)}. \tag{2.3}$$

The properties of the vector space $\mathbb{L}_2(P)$ are analogous to $\mathbb{L}_1(P)$ [16]. If $L(\lambda) = \lambda X + Y$ is in $\mathbb{L}_2(P)$ then $L(\lambda) = \lambda X^{\mathcal{B}} + Y^{\mathcal{B}}$ belongs to $\mathbb{L}_1(P)$, where the superscript $^{\mathcal{B}}$ represents blockwise transpose[2]. This connection means the action of $L(\lambda) \in \mathbb{L}_2(P)$ is characterized by a *row shift sum* operator, denoted by $\boxplus\!\!\!\downarrow$,

$$X \boxplus\!\!\!\downarrow Y = \left(X^{\mathcal{B}} \boxplus\!\!\!\rightarrow Y^{\mathcal{B}}\right)^{\mathcal{B}} = M^{\mathcal{B}}X + \begin{bmatrix} \mathbf{0}^T \\ Y \end{bmatrix}.$$

2.1. Extending the results to general polynomial bases. Many of the derivations in [20] are specifically for $P(\lambda)$ expressed in a monomial basis, though the lemmas and theorems can be generalized to any polynomial basis. One approach to generalize [20] is to use the change of basis matrix S such that $\Lambda(\lambda) = S[\lambda^{k-1}, \dots, \lambda, 1]^T$ and to define the mapping (see also [7])

$$\mathcal{C}\left(\hat{L}(\lambda)\right) = \hat{L}(\lambda)(S^{-1} \otimes I_n) = L(\lambda), \tag{2.4}$$

where $\hat{L}(\lambda)$ is a pencil in $\mathbb{L}_1(P)$ for the matrix polynomial $P(\lambda)$ expressed in the monomial basis. In particular, the strong linearization theorem holds for any polynomial basis.

THEOREM 2.1 (Strong Linearization Theorem). *Let $P(\lambda)$ be a regular matrix polynomial (expressed in any polynomial basis), and let $L(\lambda) \in \mathbb{L}_1(P)$. Then the following statements are equivalent:*

[1]Non-monomial bases are mainly of interest when working with numerical algorithms over some subfield of $\mathbb{C}$. For the sake of completeness, we note that in order to define the Chebyshev basis the field characteristics must be different than 2.

[2]If $X = (X_{ij})_{1\le i,j\le k}$, $X_{ij} \in \mathbb{F}^{n\times n}$, then $X^{\mathcal{B}} = (X_{ji})_{1\le i,j\le k}$.

1. $L(\lambda)$ is a linearization for $P(\lambda)$,
2. $L(\lambda)$ is a regular pencil,
3. $L(\lambda)$ is a strong linearization for $P(\lambda)$.

Proof. It is a corollary of [20, Theorem 4.3]. In fact, the mapping $\mathcal{C}$ in (2.4) is a strict equivalence between $\mathbb{L}_1(P)$ expressed in the monomial basis and $\mathbb{L}_1(P)$ expressed in another polynomial basis. Therefore, $L(\lambda)$ has one of the three properties if and only if $\hat{L}(\lambda)$ also does, and the properties are equivalent for $\hat{L}(\lambda)$ because they are for $L(\lambda)$. ◻

This strict equivalence can be used to generalize many properties of $\mathbb{L}_1(P)$, $\mathbb{L}_2(P)$, and $\mathbb{DL}(P)$; however, our approach based on bivariate polynomials allows for more concise derivations.

3. Recasting to bivariate matrix polynomials. A block matrix $X \in \mathbb{F}^{nk\times nh}$ with $n\times n$ blocks can provide the coefficients for a bivariate matrix polynomial of degree $h-1$ in x and $k-1$ in y. Let $\phi : \mathbb{F}^{nk\times nh} \to \mathbb{F}_{h-1}[x] \times \mathbb{F}_{k-1}[y]$ be the mapping defined by

$$\phi \;:\; X = \begin{bmatrix} X_{11} & \dots & X_{1h} \\ \vdots & \ddots & \vdots \\ X_{k1} & \dots & X_{kh} \end{bmatrix},\; X_{ij} \in \mathbb{F}^{n\times n} \mapsto F(x,y) = \sum_{i=0}^{k-1}\sum_{j=0}^{h-1} X_{k-i,k-j}\phi_i(y)\phi_j(x).$$

Equivalently, we may define the map as follows:

$$\phi \;:\; X = \begin{bmatrix} X_{11} & \dots & X_{1h} \\ \vdots & \ddots & \vdots \\ X_{k1} & \dots & X_{kh} \end{bmatrix} \mapsto F(x,y) = \begin{bmatrix} \phi_{k-1}(y)I & \cdots & \phi_0(y)I \end{bmatrix} X \begin{bmatrix} \phi_{h-1}(x)I \\ \vdots \\ \phi_0(x)I \end{bmatrix}.$$

Usually, and unless otherwise speficified, we will apply the map ϕ to square block matrices, i.e., $h = k$.

We recall that a regular (matrix) polynomial $P(\lambda)$ expressed in a degree-graded basis has an infinite eigenvalue if its leading matrix coefficient is singular. In order to correctly take care of infinite eigenvalues we write $P(\lambda) = \sum_{i=0}^{g} A_i\phi_i(\lambda)$, where the integer $g \geq k$ is called the *grade* [22]. If the grade of $P(\lambda)$ is larger than the degree then $P(\lambda)$ has at least one infinite eigenvalue. Usually, and unless stated otherwise, the grade is equal to the degree.

It is easy to show that the mapping ϕ is a bijection between $h\times k$ block matrices with $n\times n$ blocks and $n\times n$ bivariate matrix polynomials of grade $h-1$ in x and grade $k-1$ in y. Even more, ϕ is an isomorphism preserving the group additive structure. We omit the trivial proof.

Many matrix operations can be interpreted as functional operations via the above described duality between block matrices and their continuous analogues. Bivariate matrix polynomials allow us to interpret many matrix operations in terms of functional operations. In many instances, existing proofs in the theory of linearizations of matrix polynomials can be simplified, and throughout the paper we will often exploit this parallelism. We summarize some computation rules in Table 3.1. We hope the table will prove useful not only in this paper, but also for future work. All the rules are valid for any basis and for any field $\mathbb{F}$, except the last row that assumes $\mathbb{F} = \mathbb{C}$.

Other computational rules exist when the basis has additional properties. We

TABLE 3.1
Correspondence between operations in the matrix and the bivariate polynomial viewpoints.

Block matrix operation	Bivariate polynomial operation
Block matrix X	Bivariate polynomial $F(x,y)$
$X \mapsto XM$	$F(x,y) \mapsto F(x,y)x$
$X \mapsto M^{\mathcal{B}}X$	$F(x,y) \mapsto yF(x,y)$
$X(\Lambda(\lambda) \otimes I)$	Evaluation at $x = \lambda$: $F(\lambda, y)$
$X(\Lambda(\lambda) \otimes v)$	$F(\lambda, y)v$
$(\Lambda^T(\mu) \otimes w^T)X$	$w^T F(x,\mu)$
$(\Lambda^T(\mu) \otimes w^T)X(\Lambda(\lambda) \otimes v)$	$w^T F(\lambda,\mu)v$
$X \mapsto X^{\mathcal{B}}$	$F(x,y) \mapsto F(y,x)$
$X \mapsto X^T$	$F(x,y) \mapsto F^T(y,x)$
$X \mapsto X^*$	$F(x,y) \mapsto F^*(y,x)$

give some examples in Table 3.2, in which

$$\Sigma = \begin{bmatrix} \ddots & & & \\ & I & & \\ & & -I & \\ & & & I \end{bmatrix}, \quad R = \begin{bmatrix} & & I \\ & \iddots & \\ I & & \end{bmatrix}, \tag{3.1}$$

and we say that a polynomial basis is alternating if $\phi_i(x)$ is even (odd) when i is even (odd).

TABLE 3.2
Correspondence when the polynomial basis is alternating or the monomial basis.

Type of basis	Block matrix operation	Bivariate polynomial operation
Alternating	$X \mapsto \Sigma X$	$F(x,y) \mapsto F(x,-y)$
Alternating	$X \mapsto X\Sigma$	$F(x,y) \mapsto F(-x,y)$
Monomials	$X \mapsto RX$	$F(x,y) \mapsto y^{k-1}F(x,y^{-1})$
Monomials	$X \mapsto XR$	$F(x,y) \mapsto x^{h-1}F(x^{-1},y)$

As seen in Table 3.1, the matrix M in (2.1) is such that the bivariate matrix polynomial corresponding to the coefficients XM is $F(x,y)x$, i.e., M applied on the right of X represents multiplication of $F(x,y)$ by x. This gives an equivalent definition for the column shift sum operator: if the block matrices X and Y are the coefficients for $F(x,y)$ and $G(x,y)$ then the coefficients of $H(x,y)$ are Z, where

$$Z = X \boxplus\!\!\rightarrow Y, \qquad H(x,y) = F(x,y)x + G(x,y).$$

Therefore, in terms of bivariate matrix polynomials we can define $\mathbb{L}_1(P)$ as

$$\mathbb{L}_1(P) = \left\{ L(\lambda) = \lambda X + Y : F(x,y)x + G(x,y) = v(y)P(x), v \in \Pi_{h-1}(\mathbb{F}) \right\},$$

where $\Pi_{h-1}(\mathbb{F})$ is the space of polynomials in $\mathbb{F}[y]$ of degree $\leq h-1$.

Regarding the space $\mathbb{L}_2(P)$, the coefficient matrix $M^{\mathcal{B}}X$ corresponds to the bivariate matrix polynomial $yF(x,y)$, i.e., $M^{\mathcal{B}}$ applied on the left of X represents multiplication of $F(x,y)$ by y. Hence, we can define $\mathbb{L}_2(P)$ as

$$\mathbb{L}_2(P) = \left\{ L(\lambda) = \lambda X + Y : yF(x,y) + G(x,y) = P(y)w(x), w \in \Pi_{k-1}(\mathbb{F}) \right\}.$$

The space $\mathbb{DL}(P)$ is the intersection of $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. It is an important vector space because it contains block symmetric linearizations. A pencil $L(\lambda) = \lambda X + Y$ belongs to $\mathbb{DL}(P)$ with ansätze $v(y)$ and $w(x)$ if the following $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ conditions are satisfied:

$$F(x,y)x + G(x,y) = v(y)P(x), \qquad yF(x,y) + G(x,y) = P(y)w(x). \tag{3.2}$$

It appears that $v(y)$ and $w(x)$ could be chosen independently; however, if we substitute $y = x$ into (3.2) we obtain the *compatibility condition*

$$v(x)P(x) = F(x,x)x + G(x,x) = xF(x,x) + G(x,x) = P(x)w(x)$$

and hence, $v = w$ as elements of $\Pi_{k-1}(\mathbb{F})$ since $P(x)(v(x) - w(x))$ is the zero matrix. This shows the double ansatz space is actually a single ansatz space; a fact that required two quite technical proofs in [20, Prop. 5.2, Thm. 5.3].

The bivariate matrix polynomials $F(x,y)$ and $G(x,y)$ are uniquely defined by the ansatz $v(x)$ since they satisfy the explicit formulas

$$yF(x,y) - F(x,y)x = P(y)v(x) - v(y)P(x), \tag{3.3}$$

$$yG(x,y) - G(x,y)x = yv(y)P(x) - P(y)v(x)x. \tag{3.4}$$

In other words, there is an isomorphism between $\Pi_k(\mathbb{F})$ and $\mathbb{DL}(P)$. It also follows from (3.3) and (3.4) that $F(x,y) = F(y,x)$ and $G(x,y) = G(y,x)$. This shows that all the pencils in $\mathbb{DL}(P)$ are block symmetric. Furthermore, if $F(x,y)$ and $G(x,y)$ are symmetric and satisfy $F(x,y)x + G(x,y) = P(x)v(y)$ then we also have $F(y,x)x + G(y,x) = P(x)v(y)$, and by swapping x and y we obtain the $\mathbb{L}_2(P)$ condition, $yF(x,y) + G(x,y) = P(y)v(x)$. This shows all block symmetric pencils in $\mathbb{L}_1(P)$ belong to $\mathbb{L}_2(P)$ and hence, also belong to $\mathbb{DL}(P)$. Thus, $\mathbb{DL}(P)$ is the space of block symmetric pencils in $\mathbb{L}_1(P)$ [16, Thm. 3.4].

REMARK 3.1. *Although in this paper we do not consider singular matrix polynomials, we note that the analysis of this section still holds even if we drop the assumption that $P(x)$ is regular. We only need to assume $P(x) \not\equiv 0$ in our proof that $\mathbb{DL}(\mathbb{P})$ is in fact a single ansatz space. This is no loss of generality, since $\mathbb{DL}(0) = \{0\}$.*

4. Eigenvalue exclusion theorem and Bézoutians. The eigenvalue exclusion theorem [20, Thm. 6.9] shows that if $L(\lambda) \in \mathbb{DL}(P)$ with ansatz $v \in \Pi_{k-1}(\mathbb{F})$ then $L(\lambda)$ is a linearization for the matrix polynomial $P(\lambda)$ if and only if $v(\lambda)I_n$ and $P(\lambda)$ do not share an eigenvalue. This theorem is important because, generically, $v(\lambda)I_n$ and $P(\lambda)$ do not share eigenvalues and almost all choices for $v \in \Pi_{k-1}(\mathbb{F})$ correspond to linearizations in $\mathbb{DL}(P)$ for $P(\lambda)$.

THEOREM 4.1 (Eigenvalue Exclusion Theorem). *Suppose that $P(\lambda)$ is a regular matrix polynomial of degree k and $L(\lambda)$ is in $\mathbb{DL}(P)$ with a nonzero ansatz polynomial $v(\lambda)$. Then, $L(\lambda)$ is a linearization for $P(\lambda)$ if and only if $v(\lambda)I_n$ (with grade $k-1$) and $P(\lambda)$ do not share an eigenvalue.*

We note that the last statement also includes infinite eigenvalues. In the following we will observe that any $\mathbb{DL}(P)$ pencil is a (generalized) Bézout matrix and expand on this theme. This observation tremendously simplifies the proof of Theorem 4.1 and the connection with the classical theory of Bézoutian (for the scalar case) and the Lerer–Tismenetsky Bézoutian (for the matrix case) allows us to further our understanding of the $\mathbb{DL}(P)$ vector space, and leads to a new vector space of linearizations. We first recall the definition of a Bézout matrix and Bézoutian function for scalar polynomials [5, p. 277], [6, sec. 2.9].

DEFINITION 4.2 (Bézout matrix and Bézoutian function). *Let $p_1(x)$ and $p_2(x)$ be scalar polynomials*

$$p_1(x) = \sum_{i=0}^{k} a_i \phi_i(x), \qquad p_2(x) = \sum_{i=0}^{k} c_i \phi_i(x)$$

(a_k and c_k can be zero, i.e., we regard $p_1(x)$ and $p_2(x)$ as polynomials of grade k), then the Bézoutian function associated with $p_1(x)$ and $p_2(x)$ is the bivariate function

$$\mathcal{B}(p_1, p_2) = \frac{p_1(y)p_2(x) - p_2(y)p_1(x)}{x - y} = \sum_{i,j=1}^{k} b_{ij} \phi_{k-i}(y) \phi_{k-j}(x).$$

The $k \times k$ Bézout matrix associated to $p_1(x)$ and $p_2(x)$ is defined via the coefficients of the Bézoutian function

$$B(p_1, p_2) = (b_{ij})_{1 \le i,j \le k}.$$

Here are some standard properties of a Bézoutian function and Bézout matrix:

1. The Bézoutian function is skew-symmetric with respect to its polynomial arguments: $\mathcal{B}(p_1, p_2) = -\mathcal{B}(p_2, p_1)$.
2. $\mathcal{B}(p_1, p_2)$ is bilinear with respect to its polynomial arguments.
3. $B(p_1, p_2)$ is nonsingular if and only if p_1 and p_2 have no common roots.
4. $B(p_1, p_2)$ is a symmetric matrix.

Property 3 holds for polynomials whose coefficients lie in *any* field $\mathbb{F}$, provided that the common roots are sought after in the algebraic closure of $\mathbb{F}$ and roots at infinity are included. Note in fact that the dimension of the Bézout matrix depends on the formal choice of the grade of p_1 and p_2. Unusual choices of the grade are not completely artificial: for example, they may arise when evaluating a bivariate polynomial along $x = x_0$ forming a univariate polynomial [25]. Moreover, it is important to be aware that common roots at infinity make the Bézout matrix singular.

EXAMPLE 4.3. *Consider the finite field $F_2 = \{0, 1\}$ and let $p_1 = x^2$ and $p_2 = x+1$, whose finite roots are counting multiplicity $\{0, 0\}$ and $\{1\}$, respectively. The Bézout function is $x + y + xy$. If p_1 and p_2 are seen as grade 2, the Bézout matrix (in the monomial basis) is $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, which is nonsingular and has a determinant of 1. This is expected as p_1 and p_2 have no shared root. If p_1 and p_2 are seen as grade 3 the Bézout matrix becomes $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, whose kernel is spanned by $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$. Note indeed that if the grade is 3 then the roots are, respectively, $\{\infty, 0, 0\}$ and $\{\infty, \infty, 1\}$, so p_1 and p_2 share a root at ∞.*

To highlight the connection with the classic Bézout matrix we first consider scalar polynomials and show that the eigenvalue exclusion theorem immediately follows from the connection with Bézoutians.

Proof. [Proof of Theorem 4.1 for $n = 1$] Let $p(\lambda)$ be a scalar polynomial of degree (and grade) k and $v(\lambda)$ a scalar polynomial of degree $\le k - 1$. We first solve the relations in (3.3) and (3.4) to obtain

$$F(x,y) = \frac{p(y)v(x) - v(y)p(x)}{x-y}, \qquad G(x,y) = \frac{yv(y)p(x) - p(y)v(x)x}{x-y}$$

and thus, by Definition 4.2, $F(x,y) = \mathcal{B}(v,p)$ $G(x,y) = \mathcal{B}(p,vx)$ Moreover, $\mathcal{B}$ is skew-symmetric and bilinear with respect to its polynomial arguments so we have

$$L(\lambda) = \lambda X + Y = \lambda B(v,p) + B(p,xv) = -\lambda B(p,v) + B(p,xv) = B(p,(x-\lambda)v). \quad (4.1)$$

Since B is a Bézout matrix, $\det(L(\lambda)) = \det(B(p,(x-\lambda)v)) = 0$ for all λ if and only if p and v share a root. Finally, by Theorem 2.1, $L(\lambda)$ is a linearization for $p(\lambda)$ if and only if p and v do not share a root. $\square$

An alternative (more algebraic) argument is to note that p and $(x-\lambda)v$ are polynomials in x whose coefficients lie in the field of fractions $\mathbb{F}(\lambda)$. Since p has coefficients in the subfield $\mathbb{F} \subset \mathbb{F}(\lambda)$, its roots lie in the algebraic closure of $\mathbb{F}$, denoted by $\overline{\mathbb{F}}$. The factorization $(x-\lambda)v$ similarly reveals that this polynomial has one root at λ, while all the others lie in $\overline{\mathbb{F}} \cup \{\infty\}$. Therefore, P and $(x-\lambda)$ share a root in the closure of $\mathbb{F}(\lambda)$ if and only if P and v share a root in $\mathbb{F}$. Our proof of the eigenvalue exclusion theorem is purely algebraic and holds without any assumption on the field $\mathbb{F}$. However, as noted by Mehl [23], if $\mathbb{F}$ is finite it could happen that no pencil in $\mathbb{DL}$ is a linearization, because there are only finitely many choices available for the ansatz polynomial v. Although this approach is extendable to any field, for simplicity of exposition we assume for the rest of this section that the underlying field is $\mathbb{C}$.

A natural question at this point is whether this approach generalizes to the matrix case ($n > 1$). An appropriate generalization of the scalar Bézout matrix should:

- Depend on two matrix polynomials P_1 and P_2;
- Have nontrivial kernel if and only if P_1 and P_2 share an eigenvalue and the corresponding eigenvector (note that for scalar polynomials the only possible eigenvector is 1, up to multiplicative scaling).

The following examples show that the most straightforward ideas fail to satisfy the second property above.

Example 4.4. *Note first that the most naïve idea, i.e.,* $\frac{P_2(x)P_1(y) - P_1(x)P_2(y)}{x-y}$*, is generally not even a matrix polynomial.*

Almost as straightforward is the generalization $\frac{P_1(y)P_2(x) - P_1(x)P_2(y)}{x-y}$*, which is indeed a bivariate matrix polynomial. However, consider the associated Bézout block matrix. Let us check that it does not satisfy the property of being singular if and only if P_1 and P_2 have a shared eigenpair by providing two examples over the field $\mathbb{Q}$ and in the monomial basis. Consider first* $P_1 = \begin{bmatrix} x & 0 \\ 0 & x-1 \end{bmatrix}$ *and* $P_2 = \begin{bmatrix} x-6 & -1 \\ 12 & x+1 \end{bmatrix}$*. P_1 and P_2 have disjoint spectra. The corresponding Bézout matrix is* $\begin{bmatrix} 6 & 1 \\ -12 & -2 \end{bmatrix}$*, which is singular. Conversely, let* $P_1 = \begin{bmatrix} x & 1 \\ 0 & x \end{bmatrix}$ *and* $P_2 = \begin{bmatrix} 0 & x \\ x & 1 \end{bmatrix}$*. Here, P_1 and P_2 share the eigenpair* $\{0, \begin{bmatrix} 1 & 0 \end{bmatrix}^T\}$*, but the corresponding Bézout matrix is* $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$*, which is nonsingular.*

Fortunately, an extension of the Bézoutian to the matrix case was studied in the 1980s by Lerer, Tismenetsky, and others, see, e.g., [3, 18, 19] and the references therein. It turns out that it provides exactly the generalization that we need.

Definition 4.5. *For $n \times n$ regular matrix polynomials $P_1(x)$ and $P_2(x)$ of grade k, the associated Bézoutian function $\mathcal{B}_{M_2,M_1}$ is defined by [3, 18]*

$$\mathcal{B}_{M_2,M_1}(P_1,P_2) = \frac{M_2(y)P_2(x) - M_1(y)P_1(x)}{x-y} = \sum_{i,j=1}^{\ell,k} B_{ij}\phi_{\ell-i}(y)\phi_{k-j}(x), \tag{4.2}$$

where $M_1(x)$ and $M_2(x)$ are regular matrix polynomials, ℓ is the maximal degree of $M_1(x)$ and $M_2(x)$, and $M_1(x)P_1(x) = M_2(x)P_2(x)$, [9, Ch. 9]. The $n\ell \times nk$ Bézout block matrix is defined by $B_{M_2,M_1}(P_1,P_2) = (B_{ij})_{1\le i\le \ell, 1\le j\le k}$.

Note that the Lerer–Tismenetsky Bézoutian function and the corresponding Bézout block matrix are not unique as there are many possible choices of M_1 and M_2. Indeed, the matrix B does not even need to be square.

Example 4.6. *Let $P_1 = \begin{bmatrix} x & 0 \\ 0 & x-1 \end{bmatrix}$ and $P_2 = \begin{bmatrix} x-6 & -1 \\ 12 & x+1 \end{bmatrix}$ and select[3] $M_1 = \begin{bmatrix} x^2-3x+6 & x \\ 14x-12 & x^2+2x \end{bmatrix}$ and $M_2 = \begin{bmatrix} x^2+3x & 2x \\ 2x & x^2 \end{bmatrix}$. It can be verified that $M_1P_1 = M_2P_2$. The associated Lerer–Tismenetsky Bézout matrix is*

$$\begin{bmatrix} 6 & 1 \\ -12 & -2 \\ -6 & 0 \\ -12 & 0 \end{bmatrix}$$

and has a trivial kernel.

Example 4.7. *Let $P_1 = \begin{bmatrix} x & 1 \\ 0 & x \end{bmatrix}$, $P_2 = \begin{bmatrix} 0 & x \\ x & 1 \end{bmatrix}$ and select $M_1 = P_1$ and $M_2 = P_1F$, where $F = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. The Lerer–Tismenetsky Bézout matrix is $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Its kernel has dimension 2 because P_1 and P_2 only share the eigenvalue 0 and the associated Jordan chain $\begin{bmatrix} 1 & 0 \end{bmatrix}^T, \begin{bmatrix} 0 & -1 \end{bmatrix}^T$.*

When $P_1(x)$ and $P_2(x)$ commute, i.e., $P_2(x)P_1(x) = P_1(x)P_2(x)$, the natural choice of M_1 and M_2 are $M_1 = P_2$ and $M_2 = P_1$ and we write $\mathcal{B}(P_1,P_2) := \mathcal{B}_{P_1,P_2}(P_1,P_2)$. In this case the Lerer–Tismenetsky Bézout matrix is square and of size $nk \times nk$. Here are some important properties of the Lerer–Tismenetsky Bézout matrix:

1. The Bézoutian function is skew-symmetric with respect to its arguments: $\mathcal{B}_{M_2,M_1}(P_1,P_2) = -\mathcal{B}_{M_1,M_2}(P_2,P_1)$,
2. $\mathcal{B}(P_1,P_2)$ is bilinear with respect to its polynomial arguments, i.e., $\mathcal{B}(aP_1 + bP_2, P_3) = a\mathcal{B}(P_1,P_3) + b\mathcal{B}(P_2,P_3)$ if P_1, P_2 both commute with P_3,
3. The kernel of the Bézout block matrix is

$$\ker B(P_1,P_2) = \operatorname{Im} \begin{bmatrix} X_F\phi_{k-1}(T_F) \\ \vdots \\ X_F\phi_0(T_F) \end{bmatrix} \oplus \operatorname{Im} \begin{bmatrix} X_\infty\phi_0(T_\infty) \\ \vdots \\ X_\infty\phi_{k-1}(T_\infty) \end{bmatrix}, \tag{4.3}$$

and does not depend on the choice of M_1 and M_2. This was proved (in the monomial basis) in [18, Thm. 1.1]. Equation (4.3) holds for any polynomial

[3] M_1 and M_2 are of minimal degree as there are no M_1 and M_2 of degree 0 or 1 such that $M_1P_1 = M_2P_2$ exists.

basis: it can be obtained from that theorem via a congruence transformation involving the mapping $\mathcal{C}$ in (2.4). Here (X_F, T_F), (X_∞, T_∞) are the greatest common restrictions [9, Ch. 9] of the finite and infinite Jordan pairs [9, Ch.1, Ch. 7] of $P_1(x)$ and $P_2(x)$. The infinite Jordan pairs are defined regarding both polynomials as grade k.

4. If for any x and y we have $P_1(y)P_2(x) = P_2(x)P_1(y)$, then $B(P_1, P_2)$ is a (block) symmetric matrix. Note that the hypothesis is stronger than $P_1(x)P_2(x) = P_2(x)P_1(x)$, but it is always satisfied when $P_2(x) = v(x)I$.

The following lemma shows that, as in the scalar case, property 3 is the eigenvalue exclusion theorem in disguise.

Lemma 4.8. *The greatest common restriction of the (finite and infinite) Jordan pairs of the regular matrix polynomials P_1 and P_2 is nonempty if and only if P_1 and P_2 share both an eigenvalue and the corresponding eigenvector.*

Proof. Suppose that the two matrix polynomials have only finite eigenvalues. We denote by (X_1, J_1) (resp., (X_2, J_2)) a Jordan pair of P_1 (resp., P_2). Observe that a greatest common restriction is nonempty if and only if there exists at least one nonempty common restriction. First assume there exist v and x_0 such that $P_1(x_0)v = P_2(x_0)v = 0$. Up to a similarity on the two Jordan pairs (which is without loss of generality, see [9, p. 204]) we have $X_1S_1e_1 = X_2S_2e_1 = v$, $J_1S_1e_1 = S_1e_1x_0$, and $J_2S_2e_1 = S_2e_1x_0$, where S_1 and S_2 are two similarity matrices. This shows that (v, x_0) is a common restriction [9, p. 204, p.235] of the Jordan pairs of P_1 and P_2. Conversely, let (X, J) be a common restriction with J in Jordan form. We have the four equations $X_1S_1 = X$, $X = X_2S_2$, $J_1S_1 = S_1J$, and $J_2S_2 = S_2J$ for some full column rank matrices S_1 and S_2. Letting $v := Xe_1, x_0 := e_1^T Je_1$, it is easy to check that (v, x_0) is also a common restriction, and that $X_1S_1e_1 = v = X_2S_2e_1$, $J_1S_1e_1 = S_1e_1x_0$, and $J_2S_2e_1 = S_2e_1x_0$. From [9, eq. 1.64][4], it follows that $P_1(x_0)v = P_2(x_0)v = 0$.

The assumption that all the eigenvalues are finite can be easily removed (although complicating the notation appears unavoidable). In the argument above replace every Jordan pair (X, J) with a *decomposable pair* [9, pp. 188–191] of the form $[X_F, X_\infty]$ and $J_F \oplus J_\infty$, where (X_F, J_F) is a finite Jordan pair and (X_∞, J_∞) is an infinite Jordan pair [9, Ch. 7]. As the argument is essentially the same we omit the details. □

The importance of the connection with Bézout theory is now clear. The proof of the eigenvalue exclusion theorem in the matrix polynomial case becomes immediate.

Proof. [Proof of Theorem 4.1 for $n > 1$]

Let $P_1 = P(x)$ and $P_2 = (x - \lambda)v(x)I_n$ in (4.2). Then, P_1 and P_2 commute for all x, so we take $M_1 = P_2$ and $M_2 = P_1$ and obtain

$$\begin{aligned}\mathcal{B}(P(x), (x-\lambda)v(x)I_n) &= \frac{P(y)(x-\lambda)v(x) - (y-\lambda)v(y)P(x)}{x-y} \\ &= \sum_{i,j=1}^{k} B_{ij}\phi_{k-i}(y)\phi_{k-j}(x).\end{aligned}$$

This gives the $nk \times nk$ Bézout block matrix $B(P, (x - \lambda)vI) = (B_{ij})_{1 \le i,j \le k}$. Completely analogously to the scalar case, we have

$$\mathbb{DL}(P, v) = L(\lambda) = B(P, (x-\lambda)vI) = \lambda B(v, P) + B(P, xv). \tag{4.4}$$

[4]Although strictly speaking [9, eq. 1.64] is for a monic matrix polynomial, it is extendable in a straightforward way to a regular matrix polynomial (see also [9, Ch. 7]).

If vI_n and P share a finite eigenvalue λ_0 and $P(\lambda_0)w = 0$ for a nonzero w, then $(\lambda_0 - \lambda)v(\lambda_0)w = 0$ for all λ. Hence, the kernel of $L(\lambda) = B(P, (x-\lambda)v)$ is nonempty for all λ and $L(\lambda)$ is singular. An analogous argument holds for a shared infinite eigenvalue. Conversely, suppose $v(\lambda)I_n$ and $P(\lambda)$ have no common eigenvalues. If λ_0 is an eigenvalue of P then $(\lambda_0 - \lambda)v(\lambda_0)I$ is nonsingular unless $\lambda = \lambda_0$. Thus, if λ is not an eigenvalue for P then the common restriction is empty, which means $L(\lambda)$ is nonsingular. In other words, $L(\lambda)$ is regular and a linearization by Theorem 2.1. □

5. Barnett's theorem and "beyond $\mathbb{DL}$" linearization space. In this section we work for simplicity in the monomial basis, and we assume that the matrix polynomial $P(x) = \sum_{i=0}^{k} P_i x^i$ has an invertible leading coefficient P_k. Given a ring R, a *left ideal* L is a subset of R such that $(L, +)$ is a subgroup of $(R, +)$ and $r\ell \in L$ for any $\ell \in L$ and $r \in R$ [12, Ch. 1]. A *right ideal* is defined analogously.

Given a matrix polynomial $P(x)$ over some field $\mathbb{F}$ the set $L_P = \{Q(x) \in \mathbb{F}^{n\times n}[x] \mid Q(x) = A(x)P(x),\ A(x) \in \mathbb{F}^{n\times n}[x]\}$ is a left ideal of the ring $\mathbb{F}^{n\times n}[x]$. Similarly, $R_P = \{Q(x) \in \mathbb{F}^{n\times n}[x] \mid Q(x) = P(x)A(x),\ A(x) \in \mathbb{F}^{n\times n}[x]\}$ is a right ideal of $\mathbb{F}^{n\times n}[x]$.

A matrix polynomial of grade $k-1$ can be represented as $G(x) = \Gamma\Phi(x)$, where $\Gamma = [\Gamma_{k-1}, \Gamma_{k-2}, \ldots, \Gamma_0] \in \mathbb{F}^{n\times nk}$ are its coefficient matrices when expressed in the monomial basis and $\Phi(x) = \left[x^{k-1}I, \ldots, xI, I\right]^{\mathcal{B}}$. Let $C_P^{(1)}$ be the first companion matrix[5] of $P(x)$:

$$C_P^{(1)} = \begin{bmatrix} -P_k^{-1}P_{k-1} & -P_k^{-1}P_{k-2} & \cdots & -P_k^{-1}P_1 & -P_k^{-1}P_0 \\ I & & & & \\ & I & & & \\ & & \ddots & & \\ & & & I & 0 \end{bmatrix}.$$

A key observation is that the action of $C_P^{(1)}$ on Φ is that of the multiplication-by-xI operator in the quotient module R/L_P:

$$\begin{bmatrix} x^{k-1}I \\ x^{k-2}I \\ \vdots \\ xI \\ I \end{bmatrix} xI \equiv C_P^{(1)} \begin{bmatrix} x^{k-1}I \\ x^{k-2}I \\ \vdots \\ xI \\ I \end{bmatrix} = \begin{bmatrix} x^{k-1}I \\ x^{k-2}I \\ \vdots \\ xI \\ I \end{bmatrix} xI + \begin{bmatrix} -P_k^{-1}P(x) \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} -P_k^{-1}P(x) \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \in R_L^k.$$

Premultiplying by the coefficients Γ, we can identify the map $\Gamma \mapsto \Gamma C_P^{(1)}$ with the map $G(x) \mapsto G(x)x$ in $\mathbb{F}^{n\times n}[x]/L_P$. That is, we can write $\Gamma C_P^{(1)}\Phi = xG(x) + Q(x)$ for some $Q(x) \in L_P$. More precisely, we have $Q(x) = \Gamma_{k-1}P_k^{-1}P(x)$.

THEOREM 5.1. *Let $P(x) = \sum_{i=0}^{k} P_i x^i \in \mathbb{F}^{n\times n}[x]$ be a matrix polynomial of degree k such that P_k is invertible, and let $V(x) \in \mathbb{F}^{n\times n}[x]$ be any matrix polynomial. Then there exists a unique $Q(x)$ of grade $k-1$ such that $Q(x) \equiv V(x)$ in the quotient module $\mathbb{F}^{n\times n}[x]/L_P$, i.e., there exists a unique $A(x) \in \mathbb{F}^{n\times n}[x]$ such that $V(x) = A(x)P(x) + Q(x)$.*

[5] Some authors define the first companion matrix with minor differences in the choice of signs. Here, we make our choice for simplicity of what follows. For other polynomial bases the matrix should be replaced accordingly [4].

Moreover, there exists a unique $S(x)$ of grade $k-1$ such that $S(x) \equiv V(x)$ in the quotient module $\mathbb{F}^{n\times n}[x]/R_P$, i.e., there exists a unique $B(x) \in \mathbb{F}^{n\times n}[x]$ such that $V(x) = P(x)B(x) + S(x)$.

Proof. If $\deg V(x) < k$, then take $Q(x) = V(x)$ and $A(x) = 0$. If $\deg V \geq k$, then our task is to find $A(x)$ with $\deg A = \deg V - k$ that satisfies $M(x) = A(x)P(x) = \sum_{i=0}^{\deg V} M_i x^i$ and $M_i = \sum_{j+\ell=i}(A_j P_\ell) = V_i$ for $k \leq i \leq \deg V$. This is equivalent to solving the following block matrix equation:

$$\begin{bmatrix} A_{\deg A} & \cdots & A_0 \end{bmatrix} \begin{bmatrix} P_k & P_{k-1} & P_{k-2} & \cdots & & \\ & P_k & P_{k-1} & P_{k-2} & \cdots & \\ & & \ddots & \ddots & & \\ & & & P_k & P_{k-1} \\ & & & & P_k \end{bmatrix} = \begin{bmatrix} V_{\deg V} & \cdots & V_k \end{bmatrix}, \qquad (5.1)$$

which shows explicitly that $A(x)$ exists and is unique. This implies that exists unique $Q(x) = V(x) - A(x)P(x)$.

An analogous argument proves the existence and uniqueness of $B(x)$ and $S(x)$ such that $V(x) = P(x)B(x) + S(x)$. ▯

Thanks to the connection between $\mathbb{DL}$ and the Bézoutian, we find that [16, Theorem 4.1] is a generalization of Barnett's theorem to the matrix case. The proof that we give below is a generalization of that found in [13] for the scalar case. It is another example where the algebraic interpretation and the connection with Bézoutians simplify proofs (compare with the argument in [16]).

THEOREM 5.2 (Barnett's theorem for matrix polynomials). *Let $P(x)$ be a matrix polynomialof degree k with nonsingular leading coefficient and $v(x)$ a scalar polynomial of grade $k-1$. We have $\mathbb{DL}(P, v(x)) = \mathbb{DL}(P, 1)v(C_P^{(1)})$, where $C_P^{(1)}$ is the first companion matrix of $P(x)$.*

Proof. It is easy to verify that the following recurrence formula holds:

$$\frac{P(y)x^j(x-\lambda) - y^j(y-\lambda)P(x)}{x-y} = \frac{P(y)x^{j-1}(x-\lambda) - y^{j-1}(y-\lambda)P(x)}{x-y}x + y^{j-1}(y-\lambda)P(x).$$

Hence, we have $\mathcal{B}(P, (x-\lambda)x^j) \equiv \mathcal{B}(P, (x-\lambda)x^{j-1})x$ where the equivalence is in the quotient space $\mathbb{F}[y,\lambda]^{n\times n}[x]/L_P$. Taking into account the interpretation of the action of $C_P^{(1)}$ on the right as multiplication by x, this proves by induction the theorem when $v(C_P^{(1)})$ is a monomial of the form $(C_P^{(1)})^j$ for $0 \leq j \leq k-1$. The case of a generic $v(C_P^{(1)})$ follows by linearity of the Bézoutian. ▯

An analogous interpretation as a multiplication operator holds for the second companion matrix:

$$C_P^{(2)} = \begin{bmatrix} -P_{k-1}P_k^{-1} & I & & \\ -P_{k-2}P_k^{-1} & & I & \\ \vdots & & & \ddots \\ -P_1 P_k^{-1} & & & & I \\ -P_0 P_k^{-1} & & & & \end{bmatrix}.$$

Indeed, $C_P^{(2)}$ represents multiplication by y modulo the right ideal generated by $P(y)$. Again, it is thanks to (the second part of) Theorem 5.1 that this becomes a rigorous statement. A dual version of Barnett's theorem holds for the second companion matrix. Indeed, one has $\mathbb{DL}(P, v(x)) = v(C_P^{(2)})\mathbb{DL}(P, 1)$. The proof is analogous to the one for Theorem 5.2 and is omitted.

As soon as we interpret the two companion matrices in this way, we are implicitly defining a map ψ from block matrices to bivariate polynomials modulo $L_{P(x)}$ and $R_{P(y)}$, i.e., if $X \in \mathbb{F}^{nk\times nk}$ is a block matrix then the bivariate polynomial $\psi(X)$ is defined up to an additive term of the form $L(x,y)P(x) + P(y)R(x,y)$ where $L(x,y)$ and $R(y,x)$ are bivariate polynomials. In this setting, $\psi(X)$ is seen as an equivalence class. However, in this equivalence class there exists a unique bivariate polynomial having grade equal to $\deg P - 1$ separately in both x and y, as we now prove. (Clearly, this unique bivariate polynomial must be precisely $\phi(X)$, where ϕ is the map defined in section 3). The following theorem gives the appropriate matrix polynomial analogue to Euclidean polynomial division applied both in x and y.

THEOREM 5.3. *Let $P(z) = \sum_{i=0}^{k} P_i z^i \in \mathbb{F}^{n\times n}[z]$ be a matrix polynomial with P_k invertible, and let $F(x,y) = \sum_{i=0}^{k_1}\sum_{j=0}^{k_2} F_{ij}x^i y^j \in \mathbb{F}^{n\times n}[x,y]$ be a bivariate matrix polynomial. Then there is a unique decomposition $F(x,y) = Q(x,y) + A(x,y)P(x) + P(y)B(x,y) + P(y)C(x,y)P(x)$ such that*

(i) $Q(x,y)$, $A(x,y)$, $B(x,y)$ and $C(x,y)$ are all bivariate matrix polynomials,
(ii) $Q(x,y)$ has grade $k-1$ separately in x and y,
(iii) $A(x,y)$ has grade $k-1$ in y, and
(iv) $B(x,y)$ has grade $k-1$ in x.

Moreover, $Q(x,y)$ is determined uniquely by $P(z)$ and $F(x,y)$.

Proof. Let us first apply Theorem 5.1 taking $\mathbb{F}(y)$ as the base field. Then there exist unique $A(x,y)$ and $Q_1(x,y)$ such that $F(x,y) = A_1(x,y)P(x) + Q_1(x,y)$, where $A_1(x,y)$ and $Q_1(x,y)$ are polynomials in x. Furthermore, $\deg_x Q_1(x,y) \le k-1$. A priori, the entries of $A_1(x,y)$ and $Q_1(x,y)$ could be rational functions in y. However, a careful analysis of (5.1) shows that the coefficients of $A_1(x,y) = \sum_i A_{1,i}(y)x^i$ can be obtained by solving a block linear system, say, $Mw = v$, where v depends polynomially in y whereas M is *constant* in y. Hence, $A_1(x,y)$, and a fortiori $Q_1(x,y) = F(x,y) - A_1(x,y)P(x)$, are also polynomials in y. At this point we can apply Theorem 5.1 again to write (uniquely) $Q_1(x,y) = Q(x,y) + P(y)B(x,y)$ and $A_1(x,y) = A(x,y) + P(y)C(x,y)$, where $\deg_y Q(x,y)$ and $\deg_y A(x,y)$ are both $\le k-1$. Moreover, comparing again with (5.1), it is easy to check that it must also hold $\deg_x Q(x,y) \le k-1$ and $\deg_x B(x,y) \le k-1$. Hence, $F(x,y) = Q(x,y) + A(x,y)P(x) + P(y)B(x,y) + P(y)C(x,y)P(x)$ is the sought decomposition. □

The next example illustrates the concepts just introduced.

EXAMPLE 5.4. *Let $P(x) = Ix^2 + P_1x + P_0$ and consider the block matrix $X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$. We have $\phi(X) = Axy + By + Cx + D$. Let $Y = C_P^{(2)} X C_P^{(1)}$. Then we know that $\psi(Y) \equiv Ax^2y^2 + Bxy^2 + Cx^2y + Dxy$. In particular, we have $Ax^2y^2 + Bxy^2 + Cx^2y + Dxy \equiv -(P_1y - P_0)(Ax^2 + Bx) + Cx^2y + Dxy \equiv (-P_1Ay - P_0A + Cy)(-P_1x - P_0) + (D - P_1B)xy - P_0Bx = (P_1AP_1 + D - P_1B - CP_1)xy + (P_1AP_0 - CP_0)y + (P_0AP_1 - P_0B)x + P_0AP_0 = \phi(Y)$, by Theorem 5.3. Equivalenty we could have taken quotients directly on the bases. The argument is that $\begin{bmatrix} y^2I & yI \end{bmatrix} X \begin{bmatrix} x^2I \\ xI \end{bmatrix} \equiv \begin{bmatrix} -P_1y - P_0 & yI \end{bmatrix} X \begin{bmatrix} -P_1x - P_0 \\ xI \end{bmatrix} = \psi(Y)$, and leads to the same result. A third way of computing $Y = C_P^{(2)} X C_P^{(1)}$ is to formally apply the linear algebraic definition of matrix multiplication, and then apply the mapping ϕ as in Section 3 (forgetting about quotient spaces). One remarkable consequence of Theorem 5.3 is that these three* *approaches are all equivalent. Note that the same remarks apply to any block matrix of the form $\psi(v(C_P^{(2)})Xw(C_P^{(1)}))$, for any pair of polynomials $v(y)$ and $w(x)$.*

For this example, we have taken a monic $P(x)$ for simplicity. If its leading coefficient P_k is not the identity matrix, but still is nonsingular, explicit formulae get more complicated and involve P_k^{-1}.

5.1. Beyond $\mathbb{DL}$ space. The key messsage in Theorem 5.2 is that one can start with the pencil in $\mathbb{DL}$ associated with ansatz polynomial $v = 1$ and keep right multiplying by the first companion matrix $C_P^{(1)}$, thus obtaining all the pencils in the "canonical basis" of $\mathbb{DL}$ [20]. In the scalar case ($n = 1$) there is a bijection between pencils in $\mathbb{DL}$ and polynomials in $C_P^{(1)}$. However, the situation is quite different when $n > 1$, as the dimension of the space of polynomials in $C_P^{(1)}$ can have dimension up to kn, depending on the Jordan structure of $P(x)$.

REMARK 5.5. *For some matrix polynomials $P(x)$, the dimension of the polynomials in $C_P^{(1)}$ can be much lower than nk, although generically this upper bound is achieved. An extreme example is $P(x) = p(x)I$ for some scalar $p(x)$, as in this case the dimension achieves the lowest possible bound, which is k.*

It makes sense then to investigate further the pencils of the form $v(C_P^{(2)})\mathbb{DL}(P,1) = \mathbb{DL}(P,1)v(C_P^{(1)})$ for $\deg v > k-1$, because for a generic P they do not belong to $\mathbb{DL}$. We refer to the space of such pencils as the "beyond $\mathbb{DL}$" space of potential linearizations and write

$$\mathbb{DL}(P,1)v(C_P^{(1)}) =: \mathbb{BDL}(P,v).$$

Note that $\mathbb{DL}$ is now seen as a subspace of $\mathbb{BDL}$: if $\deg v \le k-1$, then $\mathbb{BDL}(P,v) = \mathbb{DL}(P,v)$.

An important fact is that, even if the degree of the polynomial $v(x)$ is larger than $k-1$, it still holds $\mathbb{BDL}(P,v) = v(C_P^{(2)})\mathbb{DL}(P,1) = \mathbb{DL}(P,1)v(C_P^{(1)})$. When $\deg v \le k-1$, i.e., for pencils in $\mathbb{DL}$, this is a consequence of the two equivalent versions of Barnett's theorem, but we now prove this more generally.

THEOREM 5.6. *For any polynomial $v(x)$, $v(C_P^{(2)})\mathbb{DL}(P,1) = \mathbb{DL}(P,1)v(C_P^{(1)})$.*

Proof. Since both $C_P^{(2)} - \lambda I$ and $C_P^{(1)} - \lambda I$ are strong linearizations of $P(\lambda)$, they have the same minimal polynomial $m(\lambda)$. Let $\gamma = \deg m(\lambda)$. By linearity, it suffices to check the statement for $v(x) = x^j$, $j = 0, \ldots, \gamma - 1$.

We give an argument by induction. Note first that the base case, i.e., $v(x) = x^0 = 1$, is a trivial identity. From the recurrence relation displayed in the proof of Barnett's theorem, we have that $\psi(\mathbb{DL}(P,1)(C_P^{(1)})^{j-1}) \equiv \mathcal{B}(P, x^{j-1}I) \bmod L_P$. By the inductive hypothesis we also have $\psi(\mathbb{DL}(P,1)(C_P^{(1)})^{j-1}) \equiv \psi((C_P^{(2)})^{j-1}\mathbb{DL}(P,1)) \equiv \mathcal{B}(P, y^{j-1}I) \bmod R_P$. Now, let $\delta(x,y) = \phi(\mathbb{DL}(P,1)(C_P^{(1)})^j - (C_P^{(2)})^j\mathbb{DL}(P,1))$. It must be $\delta(x,y) \equiv (x-y)\mathcal{B}(P, x^{j-1}I) + L(x,y)P(x) + P(y)R(x,y)$. But $(x-y)\mathcal{B}(P,x^{j-1}I) = P(y)x^{j-1} - y^{j-1}P(x)$, and hence, $\delta(x,y) \equiv 0 + L_1(x,y)P(x) + P(y)R_1(x,y)$. Applying Theorem 5.3, and noting that, by the definition of the mapping ϕ, $\delta(x,y)$ must have grade $k-1$ separately in x and y, we can argue that $\delta(x,y) = 0$. ∎

Clearly, an eigenvalue exclusion theorem continues to hold. Indeed, by assumption $\mathbb{DL}(P,1)$ is a linearization, because we suppose $P(x)$ has no eigenvalues at infinity. Thus, $\mathbb{BDL}(P,v)$ will be a linearization as long as $v(C_P^{(1)})$ is nonsingular, which happens precisely when $P(x)$ and $v(x)I$ do not share an eigenvalue. Nonetheless, it is less clear what properties, if any, pencils in $\mathbb{BDL}$ will inherit from pencils in $\mathbb{DL}$. Besides the theoretical interest of deriving its properties, $\mathbb{BDL}$ finds an application in the theory of the sign characteristics of structured matrix polynomials [1]. To investigate this matter, we will apply Theorem 5.1 taking $V(x) = v(x)I$.

To analyze the implications of Theorem 5.1 and Theorem 5.3, it is worth summarizing the theory that we have built so far with a commuting diagram. Let $\mathbb{BDL}(P, v) = \lambda X + Y$ and $\mathbb{DL}(P, 1) = \lambda \tilde{X} + \tilde{Y}$. Below, $F(x, y)$ (resp. $\tilde{F}(x, y)$) denotes the continuous analogue of X (resp. $\tilde{X}$).

$$
\begin{tikzcd}
\tilde{X} \arrow[rr, "\phi"] \arrow[dd, bend right, "A \mapsto v(C_P^{(2)})A"'] \arrow[dd, bend left, "A \mapsto Av(C_P^{(1)})"] & & \tilde{F}(x,y) \arrow[dl, "H(x,y) \mapsto v(x)H(x,y)"'] \arrow[dr, "H(x,y) \mapsto H(x,y)v(y)"] & \\
& \tilde{F}(x,y)v(x) \arrow[dr, "\text{quotient modulo } L_P"'] & & v(y)\tilde{F}(x,y) \arrow[dl, "\text{quotient modulo } R_P"] \\
X \arrow[rr, "\phi"] & & F(x,y) &
\end{tikzcd}
$$

An analogous diagram can be drawn for Y, $\tilde{Y}$, $G(x, y)$, and $\tilde{G}(x, y)$. The diagram above illustrates that we may work in the bivariate polynomial framework (right side of the diagram), which is often more convenient for algebraic manipulations than the matrix framework (left side). In particular, using Theorem 5.1, Theorem 5.3 and (3.2), we obtain the following relations:

$$v(y)P(x) \equiv S(y)P(x) = F(x,y)x+G(x,y), \;\; yF(x,y)+G(x,y) = P(y)Q(x) \equiv P(y)v(x) \tag{5.2}$$

From (5.2) it appears clear that a pencil in $\mathbb{BDL}$ generally has *distinct* left and right ansatz vectors, and that these ansatz vectors are now block vectors, associated with left and right ansatz matrix polynomials. For convenience of those readers who happen to be more familiar with the matrix viewpoint, we also display what we obtain by translating back (5.2):

$$X \boxplus\!\!\!\rightarrow Y = \begin{bmatrix} S_{k-1} \\ \vdots \\ S_0 \end{bmatrix} [P_k, P_{k-1}, \dots, P_0], \qquad X \underset{\downarrow}{\boxplus} Y = \begin{bmatrix} P_k \\ P_{k-1} \\ \vdots \\ P_0 \end{bmatrix} [Q_{k-1}, \dots, Q_0]. \tag{5.3}$$

Note that if $\deg v \leq k-1$ then $S(x) = Q(x) = v(x)I$ and we recover the familiar shifted sum equations for $\mathbb{DL}$.

The eigenvalue exclusion theorem continues to hold for $\mathbb{BDL}$ with a natural extension that replaces the ansatz vector v with the matrix polynomial Q (or S).

Theorem 5.7 (Eigenvalue exclusion theorem for $\mathbb{BDL}$). *$\mathbb{BDL}(P, v)$ is a strong linearization of $P(x)$ if and only if $P(x)$ and $Q(x)$ (or $S(x)$) do not share an eigenpair, where $Q(x)$, $S(x)$ are the unique matrix polynomials satisfying* (5.2).

Proof. We prove the eigenvalue exclusion theorem for P and Q, as the proof for P and S is analogous. We know that $\mathbb{BDL}(P, v)$ is a strong linearization if and only if we cannot find an eigenvalue x_0 and a nonzero vector w such that $P(x_0)w = v(x_0)w = 0$.

But $Q(x_0)w = v(x_0)w - A(x_0)P(x_0)w$. Hence, $P(x)$ and $v(x)I$ share an eigenpair if and only if $P(x)$ and $Q(x)$ do. ◻

We now show that pencils in $\mathbb{BDL}$ still are Lerer–Tismenetsky Bézoutians. It is convenient to first state a lemma and a corollary.

LEMMA 5.8. *Let $U \in \mathbb{F}^{nk\times nk}$ be an invertible block-Toeplitz upper-triangular matrix. Then $(U^{\mathcal{B}})^{-1} = (U^{-1})^{\mathcal{B}}$.*

Proof. We claim that, more generally, if U is an invertible Toeplitz upper-triangular matrix with elements in any ring with unity, and $L = U^T$, then $U^{-1} = (L^{-1})^T$. Taking $\mathbb{F}^{n\times n}$ as the base ring yields the statement. To prove the claim, recall that if L^{-1} exists then $L^{-1} = L^{\#}$, where the latter notation denotes the group inverse of L. Explicit formulae for $L^{\#}$ appeared in [11, eq. (3.4)][6]. Hence, it can be checked by direct computation that $(L^{-1})^T U = U(L^{-1})^T = I$. ◻

COROLLARY 5.9. *Let $U \in \mathbb{F}^{nk\times nk}$ be invertible and block Toeplitz upper-triangular, and $\Upsilon = \begin{bmatrix} v_1 I_n \\ \vdots \\ v_k I_n \end{bmatrix}$, $v_i \in \mathbb{F}$. Then $(U^{-1}\Upsilon)^{\mathcal{B}} = \Upsilon^{\mathcal{B}}(U^{\mathcal{B}})^{-1}$.*

Proof. Since the block elements of Υ commute with any other matrix, it suffices to apply Lemma 5.8. ◻

THEOREM 5.10. *If $Q(x), A(x), S(x), B(x)$ are defined as in Theorem 5.1 with $V(x) = v(x)I$, then $P(x)Q(x) = S(x)P(x)$ and $A(x) = B(x)$.*

Proof. Let $v(x)I - Q(x) = A(x)P(x)$ and $v(x)I - S(x) = P(x)B(x)$. Note first that $\deg A = \deg B = \deg v - k$ because by assumption the leading coefficient of $P(x)$ is not a zero divisor. The coefficients of $A(x)$ must satisfy (5.1), while block transposing (5.1) we obtain an equation that must be satisfied by the coefficients of $B(x)$. Equating term by term and using Corollary 5.9 we obtain $A(x) = B(x)$, and hence, $P(x)Q(x) - S(x)P(x) = P(x)B(x)P(x) - P(x)A(x)P(x) = 0$. ◻

Hence, it follows that $\mathbb{BDL}(P, v)$ is a Lerer–Tismenetsky Bézoutian; compare the result with (4.4).

COROLLARY 5.11. *It holds $\mathbb{BDL}(P, v) = \lambda B_{S,P}(Q, P) + B_{P,xS}(P, xQ)$.*

Once again, if $\deg v \le k-1$ then we recover $\mathbb{DL}(P, v)$ because $S = Q = vI$. More generally, we have $S(x) - Q(x) = [A(x), P(x)]$.

For the rest of this section, we assume that the underlying field $\mathbb{F}$ is a metric space; for simplicity, we focus on the case $\mathbb{F} = \mathbb{C}$. As mentioned in section 2.1, one property of a pencil in $\mathbb{DL}$ is block symmetry. It turns out that this property does not hold for pencils in $\mathbb{BDL}$. Nonetheless, an even deeper algebraic property is preserved. Since each matrix coefficient in a pencil in $\mathbb{DL}$ is a Lerer-Tismenetsky Bézout matrix, the inverses of those matrices are block Hankel – note that unless $n = 1$, the inverse of a block Hankel matrix needs not be block symmetric. This is a general result: a matrix is the inverse of a block Hankel if and only if it is a Lerer–Tismenetsky Bézout matrix [19, Corollary 3.4]. However, for completeness, we give a simple proof for the special case of our interest.

THEOREM 5.12. *Let $\lambda X + Y$ be a pencil either in $\mathbb{DL}$ or in $\mathbb{BDL}$ associated with a polynomial $P(x) \in \mathbb{C}[x]^{n\times n}$ with an invertible leading coefficient. Then, X^{-1} and Y^{-1} are both block Hankel matrices if the inverses exist.*

Proof. Assume first $P(0)$ is invertible, implying that $C_P^{(1)}$ is invertible as well. We have that $H_0 = (B(P, 1))^{-1}$ is block Hankel, as can be easily shown by induction

[6]It should be noted that if L^{-1} exists then L_{11} must be invertible too. Moreover, [11, Theorem 2] implies that [11, eq. (3.2)] is satisfied.

on k [9, Sec. 2.1]. By Barnett's theorem, $(C_P^{(2)})^j B(P,1) = B(P,1)(C_P^{(1)})^j$. Then $H_j := (C_P^{(1)})^{-j} H_0 = H_0(C_P^{(2)})^{-j}$. Taking into account the structure of $(C_P^{(1)})^{-1}$ and $(C_P^{(2)})^{-1}$, we see by induction that H_j is block Hankel. For a general v that does not share eigenvalues with P, we have that $(B(P,v))^{-1} = v(C_P^{(1)})^{-1} H_0$. Since $v(C_P^{(1)})^{-1}$ is a polynomial in $(C_P^{(1)})^{-1}$, this is a linear combination of the H_j, hence is block Hankel.

If $P(0)$ is singular consider any sequence $(P_n)_{n\in\mathbb{N}} = P(x)+E_n$ such that $\|E_n\| \to 0$ as $n \to \infty$ and $P_n(0) = P(0) + E_n$ is invertible $\forall n$ (such a sequence exists because singular matrices are nowhere dense). Since the Lerer–Tismenetsky Bézout matrix is linear in its arguments, $B(P_n, v) \to B(P, v)$. In particular, $B(P_n, v)$ is eventually invertible if and only if no root of v is an eigenvalue of P. The inverse is continuous as a matrix function, and thus $B(P,v)^{-1} = \lim_{n\to\infty} B(P_n, v)^{-1}$. We conclude by observing that the limit of a sequence of block Hankel matrices is block Hankel. ◻

Note that the theorem above implies that if λ_0 is not an eigenvalue of P then the evaluation of a linearization in $\mathbb{DL}$ or $\mathbb{BDL}$ at $\lambda = \lambda_0$ is the inverse of a block Hankel matrix.

Recall that a Hermitian matrix polynomial is a polynomial whose coefficients are all Hermitian matrices. If P is Hermitian we write $P^*(x) = P(x)$. It is often argued that block-symmetry is important because, if P was Hermitian in the first place and v has real coefficients, then $\mathbb{DL}(P,v)$ is also Hermitian. Although $\mathbb{BDL}(P,v)$ is not block-symmetric, it still is Hermitian when P is.

THEOREM 5.13. *Let $P(x) \in \mathbb{C}^{n\times n}[x]$ be a Hermitian matrix polynomial with invertible leading coefficient and $v(x) \in \mathbb{R}[x]$ a scalar polynomial with real coefficients. Then, $\mathbb{BDL}(P,v)$ is a Hermitian pencil.*

Proof. Recalling the explicit form of $\mathbb{BDL}(P,v) = \lambda X + Y$ from Corollary 5.11, we have $X = B_{S,P}(Q,P)$ and $Y = B_{P,xS}(P,xQ)$. Then $-X$ is associated with the Bézoutian function $F(x,y) = \frac{P(y)Q(x)-S(y)P(x)}{x-y}$. By definition, $S(x) = v(x)I - P(x)A(x)$. Taking the transpose conjugate of this equation, and noting that by assumption $P(x) = P^*(x)$, $v(x) = v^*(x)$, we obtain $S^*(x) = v(x)I - A^*(x)P(x)$. But since $Q(x)$ is unique by Theorem 5.1, $S^*(x) = Q(x)$. Hence, $F(x,y) = \frac{P(y)Q(x)-Q^*(y)P(x)}{x-y} = \frac{Q^*(y)P(x)-P(y)Q(x)}{y-x} = F^*(y,x)$, proving that X is Hermitian because the formula holds for any x, y.

Analogously $G(x,y) = \frac{P(y)xQ(x)-yQ^*(y)P(x)}{x-y} = \frac{yQ^*(y)P(x)-P(y)xQ(x)}{y-x} = G^*(y,x)$, allowing us to deduce that Y is also Hermitian. ◻

The theory of functions of a matrix [14] allows one to extend the definition of $\mathbb{BDL}$ to a general function f, rather than just a polynomial v, as long as f is defined on the spectrum of $C_P^{(1)}$ (for a more formal definition see [14]). One just puts $\mathbb{BDL}(P,f) := \mathbb{BDL}(P,v)$ where $v(x)$ is the interpolating polynomial such that $v(C_P^{(1)}) = f(C_P^{(1)})$.

COROLLARY 5.14. *Let $P(x) \in \mathbb{C}^{n\times n}[x]$ be a Hermitian matrix polynomial with invertible leading coefficient and $f : \mathbb{C} \to \mathbb{C}$ a function defined on the spectrum of $C_P^{(1)}$ and such that $f(x^*) = (f(x))^*$. Then $\mathbb{BDL}(P,f)$ is a Hermitian pencil.*

Proof. It suffices to observe that the properties of f and P imply that $f(C_P^{(1)}) = v(C_P^{(1)})$ with $v \in \mathbb{R}[x]$ [14, Def. 1.4]. ◻

In the monomial basis, other structures of interest have been defined, such as $*$-even, $*$-odd, T-even, T-odd (all these definition can be extended to any alternating basis, such as Chebyshev) or $*$-palindromic, $*$-antipalindromic, T-palindromic, T-antipalindromic. For $\mathbb{DL}$, analogues of Theorem 5.13 can be stated in all these cases

[21]. These properties extend to $\mathbb{BDL}$. We state and prove them for the $*$-even and the $*$-palindromic case:

THEOREM 5.15. *Assume that $P(x) = P^*(-x)$ is $*$-even and with an invertible leading coefficient, and that $f(x) = f^*(-x)$, and let* $\Sigma = \begin{bmatrix} \ddots & & & \\ & I & & \\ & & -I & \\ & & & I \end{bmatrix}$. *Then $\Sigma\mathbb{BDL}(P, f)$ is a $*$-even pencil. Furthermore, if $P(x) = x^k P^*(x^{-1})$ is $*$-palindromic and $f(x) = x^{k-1} f(x^{-1})$, and if* $R = \begin{bmatrix} & & & I \\ & & I & \\ & \iddots & & \\ I & & & \end{bmatrix}$, *then $R\mathbb{BDL}(P, f)$ is a $*$-palindromic pencil.*

Proof. The proof goes along the same lines as that of Theorem 5.13: we first use the functional viewpoint and the Bézoutian interpretation of $\mathbb{BDL}(P, v) = \lambda X + Y$ to map $\phi(X) = F(x, y) = \frac{-P(y)Q(x)+S(y)P(x)}{x-y}$ and $\phi(Y) = G(x, y) = \frac{P(y)xQ(x)-yS(y)P(x)}{x-y}$. Assume first that P $*$-even: we claim that $\lambda\Sigma X^* + \Sigma Y^* = -\lambda\Sigma X + \Sigma Y$. Indeed, note that the interpolating polynomial of $f(x)$ must also satisfy $v^*(x) = v(-x)$. Taking the tranpose conjugate of the equation $S(x) = v(x)I - P(x)B(x)$ we obtain $Q^*(x) = S(-x)$. This, together with Table 3.2, implies that $\phi(-\Sigma X) = \frac{P(-y)Q(x)-Q^*(y)P(x)}{x+y} = -\frac{Q^*(y)P^*(-x)-P^*(y)Q(x)}{x+y} = \phi(X^*\Sigma)$.

Similarly, $\phi(\Sigma Y) = \frac{P(-y)xQ(x)+yQ^*(y)P(x)}{x+y} = \frac{yQ^*(y)P^*(-x)+P^*(y)xQ(x)}{y+x} = \phi(Y^*\Sigma)$.

The case of a $*$-palindromic P is dealt with analgously and we omit the details. $\square$

Similar statements hold for other structures. We summarize them in the following table, omitting the proofs as they are completely analogous to those of Theorems 5.13 and 5.15.

TABLE 5.1
Structures of P, $\deg P = k$, and potential linearizations that are structure-preserving

Structure of P	Requirement on f	Pencil
Hermitian: $P(x) = P^*(x)$	$f(x^*) = f^*(x)$	$\mathbb{BDL}(P, f)$
skew-Hermitian: $P(x) = -P^*(x)$	$f(x^*) = f^*(x)$	$\mathbb{BDL}(P, f)$
symmetric: $P(x) = P^T(x)$	any $f(x)$	$\mathbb{BDL}(P, f)$
skew-symmetric: $P(x) = -P(x)^T$	any $f(x)$	$\mathbb{BDL}(P, f)$
*-even: $P(x) = P^*(-x)$	$f(x) = f^*(-x)$	$\Sigma\mathbb{BDL}(P, f)$
*-odd: $P(x) = -P^*(-x)$	$f(x) = f^*(-x)$	$\Sigma\mathbb{BDL}(P, f)$
T-even: $P(x) = P^T(-x)$	$f(x) = f(-x)$	$\Sigma\mathbb{BDL}(P, f)$
T-odd: $P(x) = -P^T(-x)$	$f(x) = f(-x)$	$\Sigma\mathbb{BDL}(P, f)$
*-palindromic: $P(x) = x^k P^*(x^{-1})$	$f(x) = x^{k-1} f^*(x^{-1})$	$R\mathbb{BDL}(P, f)$
*-antipalindromic: $P(x) = -x^k P^*(x^{-1})$	$f(x) = x^{k-1} f^*(x^{-1})$	$R\mathbb{BDL}(P, f)$
T-palindromic: $P(x) = P^*(x^{-1})$	$f(x) = x^{k-1} f(x^{-1})$	$R\mathbb{BDL}(P, f)$
T-antipalindromic: $P(x) = -P^T(x^{-1})$	$f(x) = x^{k-1} f(x^{-1})$	$R\mathbb{BDL}(P, f)$

With a similar technique, one may produce pencils with a structure that is *related* to that of the linearized matrix polynomial, e.g., if P is $*$-odd and $f(x) = -f^*(-x)$, then $\Sigma\mathbb{BDL}(P, f)$ will be $*$-even. To keep the paper within a reasonable length, we will not include a complete list of such variations on the theme. However, we note that generalizations of this kind are immediate to prove with the Bézoutian functional approach.

We conclude this section by giving the following result which has an application in the theory of sign characteristics [1]:

THEOREM 5.16. *Let $P(x)$ be $*$-palindromic of degree k, with nonsingular leading coefficient, and $f(x) = x^{k/2}$; if k is odd, suppose furthermore that the the square root is defined in such a way that $P(x)$ has no eigenvalues on the branch cut. Furthermore, $\mathbb{BDL}(P, f) = \lambda X + Y$. Then $Z = iRX$ is a Hermitian matrix.*

Proof. We claim that the statement is true when $P(x)$ has all distinct eigenvalues. Then it must be true in general. This follows by continuity, if we consider a sequence $(P_n)_n$ of $*$-palindromic polynomials converging to $P(x)$ and such that $P_n(x)$ has all distinct eigenvalues, none of which lie on the branch cut. Such a sequence exists because the set of palindromic matrix polynomials with distinct eigenvalues is dense, as can be seen arguing on the characteristic polynomial seen as a polynomial function of the $n^2(k + 1)$ independent real parameters.

It remains to prove the claim. Since X is the linear part of the pencil $\mathbb{BDL}(P, f)$, we get $\phi(X) = \frac{P(y)Q(x)-S(y)P(x)}{x-y}$, where $v(x)I = Q(x)+A(x)P(x) = S(x)+P(x)A(x)$ are defined as in Theorem 5.1 and $v(x)$ is the interpolating polynomial of $f(x)$ on the eigenvalues of $P(x)$. By assumption $P(x)$ has kn distinct eigenvalues. Denote by (λ_i, w_i, u_i), $i = 1, \ldots, nk$, an eigentriple, and consider the matrix in Vandermonde form V whose ith column is $V_i = \Lambda(\lambda_i) \otimes w_i$ (V is the matrix of eigenvectors of $C_P^{(1)}$); recall moreover that if (λ_i, w_i, u_i) is an eigentriple then $(1/\lambda_i^*, u_i^*, w_i^*)$ is. Observe that by definition $Q(\lambda_i)w_i = \lambda_i^{k/2} w_i$ and $u_i S(\lambda_i) = u_i \lambda_i^{k/2}$.

Our task is to prove that $RX = -X^*R$; observe that this is equivalent to $V^*RXV = -V^*X^*RV$. Using Table 3.1 and Table 3.2, we see that $V_i^*RXV_j$ is equal to the evaluation of $w_i^* \frac{y^k P(1/y)Q(x)-y^k S(1/y)P(x)}{xy-1} w_j$ at $(x = \lambda_j, y = \lambda_i^*)$. Suppose first that $\lambda_i\lambda_j^* \neq 1$. Then, using $P(\lambda_j)w_j = 0$ and $w_i^* P(1/\lambda_i^*) = 0$, we get $V_i^*RXV_j = 0$. When $\lambda_i^{-1} = \lambda_j^*$, we can evaluate the fraction using De L'Hôpital rule, and obtain $w_i^* \frac{-(\lambda_i^*)^k S(1/\lambda_i^*)P'(\lambda_j)}{\lambda_i^*} w_j = -w_i^*(\lambda_i^*)^{k/2-1}P'(\lambda_j)w_j$. A similar argument shows that $V_i^*X^*RV_j = w_i^*(\lambda_i^*)^{k/2-1}P'(\lambda_j)w_j$ when $\lambda_i\lambda_j^* = 1$.

We have thus shown that $V_i^*X^*RV_j = -V_i^*R^*XV_j$ for all (i, j), establishing the claim. $\square$

6. Conditioning of eigenvalues of $\mathbb{DL}(P)$. In [15], a conditioning analysis is carried out for the eigenvalues of the $\mathbb{DL}(P)$ pencils, which identifies situations in which the $\mathbb{DL}(P)$ linearization itself does not worsen the eigenvalue conditioning of the original matrix polynomial $P(\lambda)$ expressed in the monomial basis.

Here, we use the bivariate polynomial viewpoint to analyze the conditioning, using concise arguments and allowing for $P(\lambda)$ expressed in any polynomial basis. As shown in [27], the first-order expansion of a simple eigenvalue λ_i of $P(\lambda) + \Delta P(\lambda)$ is

$$\lambda_i = \lambda_i(P) - \frac{y_i^* \Delta P(\lambda_i) x_i}{y_i^* P'(\lambda_i) x_i} + \mathcal{O}(\|\Delta P(\lambda_i)\|^2), \tag{6.1}$$

where y_i and x_i are the left and right eigenvectors corresponding to λ_i.

When applied to a $\mathbb{DL}(P)$ pencil $L(\lambda) = \lambda X + Y$ with ansatz v, defining $\widehat{x}_i = x_i \otimes \Lambda(\lambda_i), \widehat{y}_i = y_i \otimes \Lambda(\lambda_i)$, where $\Lambda(\lambda) = [\phi_{k-1}(\lambda), \ldots, \phi_0(\lambda)]^T$ as before and noting that $L'(\lambda) = X$, (6.1) becomes

$$\lambda_i = \lambda_i(L) - \frac{\widehat{y}_i^* \Delta L(\lambda_i)\widehat{x}_i}{\widehat{y}_i^* X \widehat{x}_i} + \mathcal{O}(\|\Delta L(\lambda_i)\|^2), \quad i = 1, \ldots, nk. \tag{6.2}$$

Note from $X = B(v, P)$ and Table 3.1 that $\widehat{y}_i^* X \widehat{x}_i$ is evaluation of the $n \times n$ Bezoutian function $B(v(\lambda), P(\lambda))$ at $\lambda = \lambda_i$, followed by left and right multiplication by y_i^* and x_i. Therefore we have

$$\begin{aligned}
\widehat{y}_i^* X \widehat{x}_i &= y_i^* \left(\lim_{s,t \to \lambda_i} \frac{v(s)P(t) - P(s)v(t)}{s - t} \right) x_i \\
&= y_i^* \left(v'(\lambda_i)P(\lambda_i) - P'(\lambda_i)v(\lambda_i) \right) x_i \\
&= -y_i^* P'(\lambda_i) v(\lambda_i) x_i.
\end{aligned}$$

Here we used L'Hôpital's rule for the second equality and $P(\lambda_i)x_i = 0$ for the last.

Hence, the expansion (6.2) becomes

$$\lambda_i = \lambda_i(L) + \frac{1}{v(\lambda_i)} \frac{\widehat{y}_i^* \Delta L(\lambda_i)\widehat{x}_i}{y_i^* P'(\lambda_i) x_i} + \mathcal{O}(\|\Delta L(\lambda_i)\|^2). \tag{6.3}$$

Comparing (6.3) with (6.1) we see that the ratio between the perturbation of λ_i in the original $P(\lambda)$ and the linearization $L(\lambda)$ is

$$r_{\lambda_i} = \frac{1}{v(\lambda_i)} \frac{\|\widehat{y}_i\|_2 \|\Delta L(\lambda_i)\|_2 \|\widehat{x}_i\|_2}{\|y_i\|_2 \|\Delta P(\lambda_i)\|_2 \|x_i\|_2}. \tag{6.4}$$

Here we used the fact that equality in $|y_i^* \Delta P(\lambda_i) x_i| \le \|y_i\|_2 \|\Delta P(\lambda_i)\|_2 \|x_i\|_2$ and $|\widehat{y}_i^* \Delta L(\lambda_i) \widehat{x}_i| \le \|y_i\|_2 \|\Delta L(\lambda_i)\|_2 \|x_i\|_2$ can hold in (6.1) and (6.3), which can be verified by taking $\Delta P(\lambda) = \sigma_1 y_i x_i^*$ and $\Delta L(\lambda) = \sigma_2 \widehat{y}_i \widehat{x}_i^*$ for any scalars σ_1, σ_2.

Now recall that the absolute *condition number* of an eigenvalue of a matrix polynomial may be defined as

$$\kappa(\lambda) = \lim_{\epsilon \to 0} \sup\{|\Delta\lambda| : (P(\lambda + \Delta\lambda) + \Delta P(\lambda + \Delta\lambda))\,\hat{x} = 0, \hat{x} \neq 0, \|\Delta P(\cdot)\| \le \epsilon \|P(\cdot)\|\}. \tag{6.5}$$

Here, we are taking the norm for matrix polynomials to be $\|P(\cdot)\| = \max_{\lambda \in \mathcal{D}} \|P(\lambda)\|_2$, where $\mathcal{D}$ is the domain of interest that below we take to be the interval $[-1, 1]$. In (6.5), $\lambda + \Delta\lambda$ is the eigenvalue of $P + \Delta P$ closest to λ such that $\lim_{\epsilon \to 0} \Delta\lambda = 0$. Note that definition (6.5) is the *absolute* condition number, in contrast to the relative condition number treated in [27], in which the supremum is taken of $|\Delta\lambda|/(\epsilon|\lambda|)$, and over $\Delta P(\cdot) = \sum_{i=0}^{k} \Delta A_i \phi_i(\cdot)$ such that $\|\Delta A_i\|_2 \le \epsilon \|E_i\|$ where E_i are prescribed tolerances for the term with ϕ_i. Combining this definition with the analysis above, we can see that the ratio of the condition numbers of the eigenvalue λ for the linearization L and the original matrix polynomial P is

$$\widehat{r}_{\lambda_i} = \frac{1}{v(\lambda_i)} \frac{\|\widehat{y}_i\|_2 \|L(\cdot)\| \|\widehat{x}_i\|_2}{\|y_i\|_2 \|P(\cdot)\| \|x_i\|_2}. \tag{6.6}$$

The eigenvalue λ_i can be computed stably from the linearization $L(\lambda)$ if $\hat{r}_{\lambda_i}$ is not significantly larger than 1. Identifying conditions to guarantee $\hat{r}_{\lambda_i} = \mathcal{O}(1)$ is nontrivial and depends not only on $P(\lambda)$ and the choice of the ansatz v, but also on the value of λ_i and the choice of polynomial basis. For example, [15] considers the monomial case and shows that the coefficientwise conditioning of λ_i does not worsen much by forming $L(\lambda)$ if $\frac{\max_i \|A_i\|_2}{\max\{\|A_0\|_2,\|A_k\|_2\}}$ is not too large, where $P(\lambda)=\sum_{i=0}^k A_i\lambda^i$, and the ansatz choice is $v=\lambda^{k-1}$ if $|\lambda_i|\geq 1$ and $v=1$ if $|\lambda_i|\leq 1$.

Although it is difficult to make a general statement on when r_{λ_i} is moderate, here we show that in the practically important case where the Chebyshev basis is used and $\lambda_i \in \mathcal{D} := [-1,1]$, the conditioning ratio can be bounded by a modest polynomial in n and k, with an appropriate choice of v: $v=1$. This means that the conditioning of these eigenvalues does not worsen much by forming the linearization, and the eigenvalues can be computed in a stable manner from $L(\lambda)$.

Theorem 6.1. *Let $L(\lambda)$ be the $\mathbb{DL}(P)$ linearization with ansatz $v(x)=1$ of a matrix polynomial $P(\lambda)$ expressed in the Chebyshev basis. Then for any eigenvalue $\lambda_i\in[-1,1]$, the conditioning ratio $\widehat{r}_{\lambda_i}$ in (6.6) is bounded by*

$$\widehat{r}_{\lambda_i} \leq 16n(e-1)k^4. \tag{6.7}$$

Proof. Since the Chebyshev polynomials are all bounded by 1 on $[-1,1]$, we have $\|\hat{x}_i\|_2 = c_i\|x_i\|_2$, $\|\widehat{y}_i\|_2 = d_i\|y_i\|_2$ for some $c_i,d_i\in[1,\sqrt{k}]$. Therefore, we have

$$\widehat{r}_{\lambda_i} \leq \frac{k}{v(\lambda_i)}\frac{\|L(\cdot)\|}{\|P(\cdot)\|}. \tag{6.8}$$

We next claim that $\|L(\cdot)\|$ can be estimated as $\|L(\cdot)\| = \mathcal{O}(\|P(\cdot)\|\|v(\cdot)\|)$. To verify this it suffices to show that writing $L(\lambda)=\lambda X+Y$

$$\|X\|_2 \leq q_X(n,k)\|P(\cdot)\|\|v(\cdot)\|, \quad \|Y\|_2 \leq q_Y(n,k)\|P(\cdot)\|\|v(\cdot)\| \tag{6.9}$$

where q_X, q_Y are low-degree polynomials with modest coefficients. Let us first prove the bound for $\|X\|_2$ in (6.9) (to gain a qualitative understanding one can consult the construction of X, Y in section 7).

Recalling (4.4), X is the Bézout block matrix $B(vI,P)$, so its $(k-i,k-j)$ block is the coefficients for $T_i(y)T_j(x)$ of the function

$$\mathcal{B}(P,-vI) = \frac{-P(y)v(x)+v(y)P(x)}{x-y} := H(x,y).$$

Recall that $H(x,y)$ is an $n\times n$ bivariate matrix polynomial, and denote its (s,t) element by $H_{st}(x,y)$. For every fixed value of $y\in[-1,1]$, by [24, Lem. B.1] we have

$$|H_{st}(x,y)| \leq (e-1)k^2 \max_{x\in[-1,1]} |H_{st}(x,y)(x-y)| \leq 2(e-1)k^2\|P(\cdot)\|\|v(\cdot)\| \quad \text{for} \quad |x-y|\leq k^{-2},$$

and clearly

$$|H_{st}(x,y)| \leq 2k^2\|P(\cdot)\|\|v(\cdot)\| \quad \text{for} \quad |x-y|\geq k^{-2}.$$

Together we obtain $\max_{x\in[-1,1]}|H_{st}(x,y)| \leq 2(e-1)k^2\|P(\cdot)\|\|v(\cdot)\|$. Since this holds for every (i,j) and every fixed value of $y\in[-1,1]$ we obtain

$$\max_{x\in[-1,1],y\in[-1,1]} |H_{st}(x,y)| \leq 2(e-1)k^2\|P(\cdot)\|\|v(\cdot)\|. \tag{6.10}$$

To obtain (6.9) it remains to bound the coefficients in the representation of a degree-k bivariate polynomial $H_{st}(x,y)=\sum_{i=0}^k\sum_{j=0}^k h_{k-i,k-j}^{(st)}T_i(y)T_j(x)$. It holds

$$h_{k-i,k-j}^{(st)} = \left(\frac{2}{\pi}\right)^2 \int_{-1}^{1}\int_{-1}^{1} \frac{H_{st}(x,y)T_i(y)T_j(x)}{\sqrt{(1-x^2)(1-y^2)}}dxdy,$$

(for $i = k$ and $j = k$ the constant is $\frac{1}{\pi}$) and hence using $|T_i(x)| \le 1$ on $[-1,1]$ we obtain

$$\begin{aligned}
|h^{(st)}_{k-i,k-j}| &\le \left(\frac{2}{\pi}\right)^2 \max_{x\in[-1,1],y\in[-1,1]} |H_{st}(x,y)| \int_{-1}^{1}\int_{-1}^{1} \frac{1}{\sqrt{(1-x^2)(1-y^2)}} dxdy \\
&= 4 \max_{x\in[-1,1],y\in[-1,1]} |H_{st}(x,y)| \\
&\le 8(e-1)k^2\|P(\cdot)\|\|v(\cdot)\|,
\end{aligned}$$

where we used (6.10) for the last inequality. Since this holds for every (s,t) and (i,j) we conclude that

$$\|X\|_2 \le 8n(e-1)k^3\|P(\cdot)\|\|v(\cdot)\|$$

as required.

To bound $\|Y\|_2$ we use the fact that Y is the Bézout block matrix $B(P, -vxI)$, and by an analogous argument we obtain the bound

$$\|Y\|_2 \le 8n(e-1)k^3\|P(\cdot)\|\|v(\cdot)\|.$$

This establishes (6.9) with $q_X(n,k) = q_Y(n,k) = 8n(e-1)k^3$, and we obtain

$$\|L(\cdot)\| \le 16n(e-1)k^3\|P(\cdot)\|\|v(\cdot)\|. \tag{6.11}$$

Substituting this into (6.8) we obtain

$$r_{\lambda_i} \le \frac{k}{v(\lambda_i)}\frac{\|L(\cdot)\|}{\|P(\cdot)\|} \le \frac{k}{v(\lambda_i)}\frac{16n(e-1)k^3\|P(\cdot)\|\|v(\cdot)\|}{\|P(\cdot)\|}.$$

With the choice $v = 1$ we have $v(\lambda_i) = \|v(\cdot)\| = 1$, which yields (6.7). $\square$

Note that our discussion deals with the normwise condition number, as opposed to the coefficientwise condition number as treated in [15]. In practice, we observe that the eigenvalues of $L(\lambda)$ computed via the QZ algorithm are sometimes less accurate than those of $P(\lambda)$, obtained via QZ for the colleague linearization [10], which is normwise stable [24]. The reason appears to be that the backward error resulting from the colleague matrix has a special structure, but a precise explanation is an open problem.

7. Construction. We now describe an algorithm for computing $\mathbb{DL}$ pencils. The shift sum operation provides a means to obtain the $\mathbb{DL}$ pencil given the ansatz v. For general polynomial bases, however, the construction is not as trivial as for the monomial basis. We focus on the case where $\{\phi_i\}$ is an orthogonal polynomial basis, so that the multiplication matrix (2.2) has tridiagonal structure. Recall that $F(x,y)$ and $G(x,y)$ satisfy the formulas (3.2), (3.3) and (3.4). Hence for $L(\lambda) = \lambda X + Y \in \mathbb{DL}(P)$ with ansatz v, writing the bivariate equations in terms of their coefficient matrix expansions, we see that X and Y need to satisfy the following equations: defining $v = [v_{k-1}, \dots, v_0]^T$ to be the vector of coefficients of the ansatz,

$$S = v \otimes [A_k, A_{k-1}, \dots, A_0] \qquad \text{and} \qquad T = v^T \otimes [A_k, A_{k-1}, \dots, A_0]^{\mathcal{B}},$$

by (3.4) we have

$$\begin{bmatrix} 0 \\ Y \end{bmatrix} M - M^T \begin{bmatrix} 0 & Y \end{bmatrix} = TM - M^T S, \tag{7.1}$$

and

$$XM = S - \begin{bmatrix} 0 & Y \end{bmatrix}. \tag{7.2}$$

Note that we have used the first equation of (3.2) instead of (3.3) to obtain an equation for X because the former is simpler to solve. Now we turn to the computation of X, Y, which also explicitly shows that X, Y satisfying (7.1), (7.2) is unique[7]. We first solve (7.1) for Y. Recall that M in (2.3) is block tridiagonal, the (i,j) block being $m_{i,j}I_n$. Defining $R = TM - M^TS$ and denoting by Y_i, R_i the ith block rows of Y and R respectively, the first block row of (7.1) yields $m_{1,1}Y_1 = -R_1$, hence $Y_1 = -\frac{1}{m_{1,1}}R_1$ (note that $m_{i,i} \neq 0$ because the polynomial basis is degree-graded). The second block row of (7.1) gives $Y_1M-(m_{1,2}Y_1+m_{2,2}Y_2) = R_2$, hence $Y_2 = \frac{1}{m_{2,2}}(Y_1M-m_{1,2}Y_1-R_2)$. Similarly, from the $i(\geq 3)$th block row of (7.1) we get

$$Y_i = \frac{1}{m_{i,i}}(Y_{i-1}M - m_{i-2,i}Y_{i-2} - m_{i-1,i}Y_{i-1} - R_i),$$

so we can compute Y_i for $i = 1, 2, \ldots, n$ inductively. Once Y is obtained, X can be computed easily by (7.2). The complexity is $\mathcal{O}((nk)^2)$, noting that $Y_{i-1}M$ can be computed with $\mathcal{O}(n^2k)$ cost. In Subsection 7.1 we provide a Matlab code that computes $\mathbb{DL}(P)$ for any orthogonal polynomial basis.

If $P(\lambda)$ is expressed in the monomial basis we have (see [6, Eqn. 2.9.3] for scalar polynomials)

$$L(\lambda) = \begin{bmatrix} A_{k-1} & \dots & A_0 \\ \vdots & \iddots & \\ A_0 & & \end{bmatrix} \begin{bmatrix} \hat{v}_k I_n & \dots & \hat{v}_1 I_n \\ & \ddots & \vdots \\ & & \hat{v}_k I_n \end{bmatrix} - \begin{bmatrix} \hat{v}_{k-1} I_n & \dots & \hat{v}_0 I_n \\ \vdots & \iddots & \\ \hat{v}_0 I_n & & \end{bmatrix} \begin{bmatrix} A_k & \dots & A_1 \\ & \ddots & \vdots \\ & & A_k \end{bmatrix},$$

where $\hat{v}_i = (v_{i-1} - \lambda v_i)$. This relation can be used to obtain expressions for the block matrices X and Y. For other orthogonal bases the relation is more complicated.

Matrix polynomials expressed in the Legendre or Chebyshev basis are of practical importance, for example, for a nonlinear eigenvalue solver based on Chebyshev interpolation [8]. Following [20, Table 5.2], in Table 7.1 we depict three $\mathbb{DL}(P)$ pencils for the cubic matrix polynomial $P(\lambda) = A_3T_3(\lambda) + A_2T_2(\lambda) + A_1T_1(\lambda) + A_0T_0(\lambda)$, where $T_j(\lambda)$ is the jth Chebyshev polynomial.

7.1. Matlab code for $\mathbb{DL}(P)$. The formulae (3.3) and (3.4) can be used to construct any pencil in $\mathbb{DL}(P)$ without basis conversion, which can be numerically important [2, 25]. We provide a Matlab code that constructs pencils in $\mathbb{DL}(P)$ when the matrix polynomial is expressed in any orthogonal basis. If $P(\lambda)$ is expressed in the monomials then `a = [ones(k,1)]; b = zeros(k,1); c = zeros(k,1);` and if expressed in the Chebyshev basis then `a = [ones(k - 1,1);2]/2; b = zeros(k,1); c = ones(k,1)/2;`.

```
function [X Y] = DLP(AA,v,a,b,c)
%DLP constructs the DL pencil with ansatz vector v.
% [X,Y] = DLP(AA,v,a,b,c) returns the DL pencil lambda*X + Y
%    corresponding to the matrix polynomial with coefficients AA in an
%    orthogonal basis defined by the recurrence relations a, b, c.
```

[7] We note that (7.1) is a singular Sylvester equation, but if we force the zero structure in the first block column in $[0\ Y]$ then the solution becomes unique.

Table 7.1

Three instances of pencils in $\mathbb{DL}(P)$ and their linearization condition for the cubic matrix polynomial $P(\lambda) = A_3T_3(\lambda) + A_2T_2(\lambda) + A_1T_1(\lambda) + A_0T_0(\lambda)$, expressed in the Chebyshev basis of the first kind. These three pencils form a basis for the vector space $\mathbb{DL}(P)$.

v	$L(\lambda) \in \mathbb{DL}(P)$ for given v	Linearization condition
$\begin{bmatrix}1\\0\\0\end{bmatrix}$	$\lambda\begin{bmatrix}2A_3 & 0 & 0\\ 0 & 2A_3-2A_1 & -2A_0\\ 0 & -2A_0 & A_3-A_1\end{bmatrix} + \begin{bmatrix}A_2 & A_1-A_3 & A_0\\ A_1-A_3 & 2A_0 & A_1-A_3\\ A_0 & A_1-A_3 & A_0\end{bmatrix}$	$\det(A_0 + \frac{-A_3+A_1}{\sqrt{2}}) \neq 0$ $\det(A_0 - \frac{-A_3+A_1}{\sqrt{2}}) \neq 0$
$\begin{bmatrix}0\\1\\0\end{bmatrix}$	$\lambda\begin{bmatrix}0 & 2A_3 & 0\\ 2A_3 & 2A_2 & 2A_3\\ 0 & 2A_3 & A_2-A_0\end{bmatrix} + \begin{bmatrix}-A_3 & 0 & -A_3\\ 0 & A_1-3A_3 & A_0-A_2\\ -A_3 & A_0-A_2 & -A_3\end{bmatrix}$	$\det(-A_2+A_0) \neq 0$ $\det(A_3) \neq 0$
$\begin{bmatrix}0\\0\\1\end{bmatrix}$	$\lambda\begin{bmatrix}0 & 0 & 2A_3\\ 0 & 4A_3 & 2A_2\\ 2A_3 & 2A_2 & A_1+A_3\end{bmatrix} + \begin{bmatrix}0 & -2A_3 & 0\\ -2A_3 & -2A_2 & -2A_3\\ 0 & -2A_3 & A_0-A_2\end{bmatrix}$	$\det(A_3) \neq 0$

```
[n m] = size(AA); k=m/n-1; s=n*k;              % matrix size & degree
M = spdiags([a b c],[0 1 2],k,k+1);
M = kron(M,eye(n));                            % multiplication matrix

S = kron(v,AA);
for j=0:k-1, jj=n*j+1:n*j+n; AA(:,jj)=AA(:,jj)';end % block transpose
T = kron(v.',AA'); R=M'*S-T*M;                 % construct RHS

% The Bartels-Stewart algorithm on M'Y+YM=R
X = zeros(s); Y=X; ii=n+1:s+n; nn=1:n;         % useful indices
Y(nn,:)=R(nn,ii)/M(1); X(nn,:)=T(nn,:)/M(1);    % 1st column of X and Y
Y(nn+n,:)=(R(nn+n,ii)-M(1,n+1)*Y(nn,:)+Y(nn,:)*M(:,n+1:s+n))/M(n+1,n+1);
X(nn+n,:)=(T(nn+n,:)-Y(nn,:)-M(1,n+1)*X(nn,:))/M(n+1,n+1); % 2nd cols

for i = 3:k                                    % backwards subs
    ni=n*i; jj=ni-n+1:ni; j0=jj-2*n; j1=jj-n;   % useful indices
    M0=M(ni-2*n,ni); M1=M(ni-n,ni); m=M(ni,ni); % consts of 3-term
    Y0=Y(j0,:); Y1=Y(j1,:); X0=X(j0,:); X1=X(j1,:); % vars in 3-term
    Y(jj,:)=(R(jj,ii)-M1*Y1-M0*Y0+Y1*M(:,n+1:s+n))/m;
    X(jj,:)=(T(jj,:)-Y1-M1*X1-M0*X0)/m;         % use Y to solve for X
end
```

Acknowledgements. We wish to thank Nick Higham, Françoise Tisseur, and Nick Trefethen for their support and insightful comments. We thank Leiba Rodman for his words of appreciation and encouragement, and we posthumously dedicate this article to him.

REFERENCES

[1] M. Al-Ammari, Y. Nakatsukasa, V. Noferini, and F. Tisseur. Structured matrix polynomials and their sign characteristic. In preparation.

[2] A. Amiraslani, R. M. Corless, and P. Lancaster. Linearization of matrix polynomials expressed in polynomial bases. *IMA J. Numer. Anal.*, 29(1):141–157, 2009.

[3] B. Anderson and E. Jury. Generalized Bezoutian and Sylvester matrices in multivariable linear control. *IEEE Trans. Autom. Control.*, 21(4):551–556, 1976.

[4] S. Barnett. *Polynomials and Linear Control Systems*. Marcel Dekker Inc., 1983.

[5] D. S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, 2nd edition, 2009.

[6] D. Bini and V. Y. Pan. *Polynomial and Matrix Computations (vol. 1): Fundamental Algorithms*. Birkhauser Verlag, 1994.

[7] F. De Terán, F. M. Dopico, and D. S. Mackey. Linearizations of singular matrix polynomials and the recovery of minimal indices. *Electron. J. Linear Algebra*, 18:371–402, 2009.

[8] C. Effenberger and D. Kressner. Chebyshev interpolation for nonlinear eigenvalue problems. *BIT*, 52:933–951, 2012.

[9] I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials*. SIAM, Philadelphia, USA, 2009. Unabridged republication of book first published by Academic Press in 1982.

[10] I. J. Good. The colleague matrix, a Chebyshev analogue of the companion matrix. *The Quarterly Journal of Mathematics*, 12(1):61–68, 1961.

[11] R. E. Hartwig and J. Shoaf. Group inverses and Drazin inverses of bidiagonal and triangular Toeplitz matrices. *J. Austral. Math. Soc.*, 24(Series A):10–34, 1977.

[12] M. Hazewinkel, N. Gubareni, and V. V. Kirichenko. *Algebras, Rings and Modules*, volume 1. Springer, 2004.

[13] G. Heinig and K. Rost. *Algebraic Methods for Toeplitz-like Matrices and Operators*. Birkhäuser Basel, 1984.

[14] N. J. Higham. *Functions of Matrices: Theory and Computation*. SIAM, Philadelphia, PA, USA, 2008.

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

[16] N. J. Higham, D. S. Mackey, N. Mackey, and F. Tisseur. Symmetric linearizations for matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 29(1):143–159, 2006.

[17] D. Kressner and J. E. Roman. Memory-efficient Arnoldi algorithms for linearizations of matrix polynomials in Chebyshev basis. *Numer. Lin. Alg. Appl.*, 21(4):569–588, 2014.

[18] L. Lerer and M. Tismenetsky. The Bezoutian and the eigenvalue-separation problem for matrix polynomials. *Integral Equations and Operator Theory*, 5(1):386–445, 1982.

[19] L. Lerer and M. Tismenetsky. Generalized Bezoutian and the inversion problem for block matrices, I. General scheme. *Integral Equations and Operator Theory*, 9(6):790–819, 1986.

[20] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Vector spaces of linearizations for matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 28:971–1004, 2005.

[21] 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(4):1029–1051, 2006.

[22] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Smith forms of palindromic matrix polynomials. *Electron. J. Linear Algebra*, 22:53–91, 2011.

[23] C. Mehl. Private communication. 2013.

[24] Y. Nakatsukasa and V. Noferini. On the stability of computing polynomial roots via confederate linearizations. Technical Report 2014.49, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2014. To appear in Math. Comp.

[25] Y. Nakatsukasa, V. Noferini, and A. Townsend. Computing the common zeros of two bivariate functions via Bézout resultants. *Numerische Mathematik*, 129(1):181–209, 2015.

[26] W. Specht. Die lage der nullstellen eines polynoms. III. *Mathematische Nachrichten*, 16(5-6):369–389, 1957.

[27] F. Tisseur. Backward error and condition of polynomial eigenvalue problems. *Linear Algebra Appl.*, 309(1):339–361, 2000.