Vector Spaces of Linearizations for Matrix Polynomials

Mackey, D. Steven and Mackey, Niloufer and Mehl, Christian and Mehrmann, Volker

2006

Manchester Institute for Mathematical Sciences

School of Mathematics

The University of Manchester

VECTOR SPACES OF LINEARIZATIONS
FOR MATRIX POLYNOMIALS*

D. STEVEN MACKEY†, NILOUFER MACKEY‡, CHRISTIAN MEHL§, AND VOLKER
MEHRMANN§

Abstract. The classical approach to investigating polynomial eigenvalue problems is linearization, where the polynomial is converted into a larger matrix pencil with the same eigenvalues. For any polynomial there are infinitely many linearizations with widely varying properties, but in practice the companion forms are typically used. However, these companion forms are not always entirely satisfactory, and linearizations with special properties may sometimes be required.

Given a matrix polynomial P, we develop a systematic approach to generating large classes of linearizations for P. We show how to simply construct two vector spaces of pencils that generalize the companion forms of P, and prove that almost all of these pencils are linearizations for P. Eigenvectors of these pencils are shown to be closely related to those of P. A distinguished subspace is then isolated, and the special properties of these pencils are investigated. These spaces of pencils provide a convenient arena in which to look for structured linearizations of structured polynomials, as well as to try to optimize the conditioning of linearizations.

Key words. matrix polynomial, matrix pencil, linearization, strong linearization, shifted sum, companion form

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

DOI. 10.1137/050628350

1. Introduction. Polynomial eigenvalue problems $P(\lambda)x = 0$, where $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ with real or complex coefficient matrices A_i, form the basis for (among many other applications) the vibration analysis of buildings, machines, and vehicles [5], [9], [21], and numerical methods for the solution of these problems are incorporated into most commercial and noncommercial software packages for structural analysis.

The classical and most widely used approach to solving polynomial eigenvalue problems is *linearization*, i.e., the conversion of $P(\lambda)x = 0$ into a larger size linear eigenvalue problem $L(\lambda)z = (\lambda X + Y)z = 0$ with the same eigenvalues, so that classical methods for linear eigenvalue problems can be pressed into service. The linearizations most commonly commissioned are the companion forms for $P(\lambda)$, one of which is

$$(1.1) \qquad L(\lambda) = \lambda \begin{bmatrix} A_k & 0 & \cdots & 0 \\ 0 & I_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_n \end{bmatrix} + \begin{bmatrix} A_{k-1} & A_{k-2} & \cdots & A_0 \\ -I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -I_n & 0 \end{bmatrix}.$$

†School of Mathematics, The University of Manchester, Sackville Street, Manchester, M60 1QD, UK (smackey@ma.man.ac.uk). The work of this author was supported by Engineering and Physical Sciences Research Council grant GR/S31693.
‡Department of Mathematics, Western Michigan University, Kalamazoo, MI 49008 (nil.mackey@wmich.edu; http://homepages.wmich.edu/~mackey).
§Institut für Mathematik, MA 4-5, Technische Universität Berlin, D-10623 Berlin, Germany (mehl@math.tu-berlin.de, mehrmann@math.tu-berlin.de; also http://www.math.tu-berlin.de/~mehl, http://www.math.tu-berlin.de/~mehrmann).

Many physical problems lead to matrix polynomials that are structured in some way; for example, the coefficient matrices may all be symmetric [9], or perhaps alternate between symmetric and skew-symmetric [15], or even have palindromic structure [12]. Such structure in the matrix polynomial often forces symmetries or constraints on the spectrum [12], [14], [15], [21] that have physical significance. Numerical methods (in a finite precision environment) that ignore this structure often destroy these qualitatively important spectral symmetries, sometimes even to the point of producing physically meaningless or uninterpretable results [21].

Unfortunately the companion form linearizations do not reflect any structure that may be present in the original polynomial, so their use for numerical computation in such situations may be problematic. It would be preferable if the structural properties of the polynomial were faithfully reflected in the linearization; a structure-preserving numerical method that leaves the qualitative properties of the spectrum intact would then be possible. Examples of such structured linearizations and their concomitant structure-preserving numerical methods can be found in [14] and [15].

An important issue for any computational problem is its conditioning, i.e., its sensitivity to small perturbations. It is known that different linearizations for a given polynomial eigenvalue problem can have very different conditioning [20], [21], so that numerical methods may produce rather different results for each linearization. It would clearly be useful to have available a large class of easily constructible linearizations from which one could always select a linearization guaranteed to be as well-conditioned as the original problem.

A further issue for linearizations concerns eigenvalues at ∞. Much of the literature on polynomial eigenvalue problems considers only polynomials whose leading coefficient matrix A_k is nonsingular (or even the identity), so the issue of infinite eigenvalues does not even arise. However, there are a number of applications, such as constraint multibody systems [2], [16], circuit simulation [3], or optical waveguide design [17], where the leading coefficient is singular. In such cases one must choose a linearization with care, since not all linearizations properly reflect the Jordan structure of the eigenvalue ∞ [13]. It has therefore been suggested [4], [10] that only *strong linearizations*, which are guaranteed to preserve the structure of infinite eigenvalues, can safely be used in such circumstances. Having a large class of linearizations that are known to also be strong linearizations would make this issue of infinite eigenvalues less of a concern in practice.

The aim of this paper is to show how to systematically generate, for any regular matrix polynomial P, large classes of linearizations that address these issues. These linearizations are easy to construct from the data in P, properly handle any infinite eigenvalues, provide a fertile source of structured linearizations for many types of structured polynomials [7], [12], and collectively constitute a well-defined arena in which to look for "optimally" conditioned linearizations [8].

After introducing some basic definitions and notation in section 2, we develop a natural generalization of the companion forms in section 3. The result is two large vector spaces of pencils for each matrix polynomial P, termed $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. Eigenvectors of any pencil from these associated vector spaces are shown to be simply related to the eigenvectors of P, thereby deepening the analogy to the companion forms. While not every pencil in these spaces is a linearization for P, we describe conditions under which these pencils are linearizations in section 4. As a consequence we can then show that almost every pencil in these spaces is in fact a strong linearization for P.

Finally, pencils in $\mathbb{L}_1(P) \cap \mathbb{L}_2(P)$ are considered in sections 5 and 6. For a polynomial P of degree k this intersection, termed $\mathbb{DL}(P)$, is shown to be a subspace of dimension k. Further properties of these special pencils are derived, including an elegant characterization of exactly which pencils in $\mathbb{DL}(P)$ are linearizations for P.

Subsequent papers [7], [8], [12] extend this work to the investigation of the conditioning of linearizations in $\mathbb{DL}(P)$ and the construction of structured linearizations for various types of structured matrix polynomials.

2. Basic definitions and notation. We study $n \times n$ matrix polynomials of the form

$$(2.1) \qquad P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i, \quad A_0, \dots, A_k \in \mathbb{F}^{n \times n}, \; A_k \neq 0,$$

where $\mathbb{F}$ denotes the field of real or complex numbers and k is the degree of P.

DEFINITION 2.1. *If $\lambda \in \mathbb{C}$ and nonzero $x \in \mathbb{C}^n$ satisfy $P(\lambda)x = 0$, then x is said to be a right eigenvector of P corresponding to the (finite) eigenvalue λ.*

Following standard usage, we will often abbreviate "right eigenvector" to just "eigenvector" when there is no ambiguity.

Our main concern is with *regular matrix polynomials*, i.e., polynomials $P(\lambda)$ such that $\det P(\lambda)$ is not identically zero for all $\lambda \in \mathbb{C}$; for such polynomials the finite eigenvalues are precisely the roots of the scalar polynomial $\det P(\lambda)$. Note, however, that some of our results also hold for singular matrix polynomials (these are studied in detail in [13], [18]).

It is also useful to allow ∞ as a possible eigenvalue of $P(\lambda)$. The technical device underlying this notion is the correspondence between the eigenvalues of P and those of the polynomial obtained from P by reversing the order of its coefficient matrices.

DEFINITION 2.2 (Reversal of matrix polynomials). *For a matrix polynomial $P(\lambda)$ of degree k as in (2.1), the* reversal *of $P(\lambda)$ is the polynomial*

$$(2.2) \qquad \mathrm{rev}\, P(\lambda) := \lambda^k P(1/\lambda) = \sum_{i=0}^{k} \lambda^i A_{k-i}.$$

Note that the nonzero finite eigenvalues of $\mathrm{rev}\, P$ are the reciprocals of those of P; the next definition shows how in this context we may also sensibly view 0 and ∞ as reciprocals.

DEFINITION 2.3 (Eigenvalue at ∞). *Let $P(\lambda)$ be a regular matrix polynomial of degree $k \geq 1$. Then $P(\lambda)$ is said to have an eigenvalue at ∞ with eigenvector x if $\mathrm{rev}\, P(\lambda)$ has the eigenvalue 0 with eigenvector x. The algebraic, geometric, and partial multiplicities of the infinite eigenvalue are defined to be the same as the corresponding multiplicities of the zero eigenvalue of $\mathrm{rev}\, P(\lambda)$.*

The classical approach to solving and investigating polynomial eigenvalue problems $P(\lambda)x = 0$ is to first perform a *linearization*, that is, to transform the given polynomial into a linear matrix pencil $L(\lambda) = \lambda X + Y$ with the same eigenvalues, and then work with this pencil. This transformation of polynomials to pencils is mediated by *unimodular* matrix polynomials, i.e., matrix polynomials $E(\lambda)$ such that $\det E(\lambda)$ is a nonzero constant, independent of λ.

DEFINITION 2.4 (Linearization [5]). *Let $P(\lambda)$ be an $n \times n$ matrix polynomial of degree k with $k \geq 1$. A pencil $L(\lambda) = \lambda X + Y$ with $X, Y \in \mathbb{F}^{kn \times kn}$ is called a linearization of $P(\lambda)$ if there exist unimodular matrix polynomials $E(\lambda), F(\lambda)$ such*

that

$$E(\lambda)L(\lambda)F(\lambda) = \left[\begin{array}{c|c} P(\lambda) & 0 \\ \hline 0 & I_{(k-1)n} \end{array} \right].$$

There are many different possibilities for linearizations, but probably the most important examples in practice have been the so-called companion forms or companion polynomials [5]. Letting

(2.3a) $$X_1 = X_2 = \mathrm{diag}(A_k, I_{(k-1)n}),$$

(2.3b) $$Y_1 = \begin{bmatrix} A_{k-1} & A_{k-2} & \cdots & A_0 \\ -I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -I_n & 0 \end{bmatrix}, \quad \text{and} \quad Y_2 = \begin{bmatrix} A_{k-1} & -I_n & \cdots & 0 \\ A_{k-2} & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & -I_n \\ A_0 & 0 & \cdots & 0 \end{bmatrix},$$

then $C_1(\lambda) = \lambda X_1 + Y_1$ and $C_2(\lambda) = \lambda X_2 + Y_2$ are, respectively, called the *first* and *second companion forms* for $P(\lambda)$ in (2.1).

The notion of linearization in Definition 2.4 has been designed mainly for matrix polynomials (2.1) with invertible leading coefficient A_k. In this case all the eigenvalues of $P(\lambda)$ are finite, and their Jordan structures (i.e., their partial multiplicities) may be recovered from *any* linearization [5]. However, the situation is somewhat different when the leading coefficient of a regular $P(\lambda)$ is singular, so that ∞ is an eigenvalue with some multiplicity $m > 0$. Although the Jordan structures of all the finite eigenvalues of P are still faithfully recovered from any linearization of P, the eigenvalue ∞ is problematic. Consider, for example, the fact that the identity matrix is a linearization for any unimodular $P(\lambda)$. Indeed, in [10] it is shown that *any* Jordan structure for the eigenvalue ∞ that is compatible with its algebraic multiplicity m can be realized by some linearization for P. Thus linearization in the sense of Definition 2.4 completely fails to reflect the Jordan structure of infinite eigenvalues.

To overcome this deficiency, a modification of Definition 2.4 was introduced in [4], and termed *strong linearization* in [10]. The correspondence between the infinite eigenvalue of a matrix polynomial P and the eigenvalue zero of $\mathrm{rev}\, P$ is the source of this strengthened definition.

DEFINITION 2.5 (Strong Linearization). *Let $P(\lambda)$ be a matrix polynomial of degree k with $k \geq 1$. If $L(\lambda)$ is a linearization for $P(\lambda)$ and $\mathrm{rev}\, L(\lambda)$ is a linearization for $\mathrm{rev}\, P(\lambda)$, then $L(\lambda)$ is said to be a strong linearization for $P(\lambda)$.*

For regular polynomials $P(\lambda)$, the additional property that $\mathrm{rev}\, L(\lambda)$ is a linearization for $\mathrm{rev}\, P(\lambda)$ ensures that the Jordan structure of the eigenvalue ∞ is preserved by strong linearizations. The first and second companion forms of any regular polynomial P have this additional property [4], and thus are always strong linearizations for P. Most of the pencils we construct in this paper will be shown to be strong linearizations.

The following notation will be used throughout the paper: $I = I_n$ is the $n \times n$ identity, $R = R_k$ denotes the $k \times k$ reverse identity, and $N = N_k$ is the standard $k \times k$ nilpotent Jordan block, i.e.,

(2.4) $$R = R_k = \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix}, \quad \text{and} \quad N = N_k = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}.$$

The vector $\begin{bmatrix} \lambda^{k-1} & \lambda^{k-2} & \cdots & \lambda & 1 \end{bmatrix}^T$ of decreasing powers of λ is denoted by Λ. We will also sometimes use Λ with an argument, so that $\Lambda(r) = \begin{bmatrix} r^{k-1} & r^{k-2} & \cdots & r & 1 \end{bmatrix}^T$. Denoting the Kronecker product by $\otimes$, the unimodular matrix polynomials

$$(2.5) \quad T(\lambda) = \begin{bmatrix} 1 & \lambda & \lambda^2 & \cdots & \lambda^{k-1} \\ & 1 & \lambda & \ddots & \vdots \\ & & 1 & \ddots & \lambda^2 \\ & & & \ddots & \lambda \\ & & & & 1 \end{bmatrix} \otimes I \quad \text{and} \quad G(\lambda) = \begin{bmatrix} 1 & & & \lambda^{k-1} \\ & \ddots & & \vdots \\ & & 1 & \lambda \\ & & & 1 \end{bmatrix} \otimes I$$

are used in several places in this paper. Observe that the last block-column of $G(\lambda)$ is $\Lambda \otimes I$, and that $T(\lambda)$ may be factored as

$$(2.6) \quad T(\lambda) = G(\lambda) \begin{bmatrix} I & \lambda I & & \\ & I & & \\ & & I & \\ & & & \ddots \\ & & & & I \end{bmatrix} \begin{bmatrix} I & & & \\ & I & \lambda I & \\ & & I & \\ & & & \ddots \\ & & & & I \end{bmatrix} \cdots \begin{bmatrix} I & & & \\ & \ddots & & \\ & & I & \lambda I \\ & & & I \\ & & & & I \end{bmatrix}.$$

3. Vector spaces of "potential" linearizations. The companion forms of a matrix polynomial $P(\lambda)$ have several nice properties that make them attractive as linearizations for P:

- they are immediately constructible from the data in P,
- eigenvectors of P are easily recovered from eigenvectors of the companion forms,
- they are always strong linearizations for P.

However, the companion forms have one significant drawback; they usually do not reflect any structure or eigenvalue symmetry that may be present in the original polynomial P. One would like to be able to draw on a source of linearizations for P that allow for the preservation of structure while sharing as many of the useful properties of companion forms as possible. To this end we introduce vector spaces of pencils that generalize the two companion forms, and analyze some of the properties these pencils have in common with the companion forms.

To motivate the definition of these spaces, let us recall the origin of the first companion form. Imitating the standard procedure for converting a system of higher order linear differential algebraic equations into a first order system (see [5]), introduce the variables $x_1 = \lambda^{k-1}x$, $x_2 = \lambda^{k-2}x, \ldots, x_{k-1} = \lambda x$, $x_k = x$, thereby transforming the $n \times n$ polynomial eigenvalue problem $P(\lambda)x = (\sum_{i=0}^{k} \lambda^i A_i)x = 0$ into

$$A_k(\lambda x_1) + A_{k-1}x_1 + A_{k-2}x_2 + \cdots + A_1 x_{k-1} + A_0 x_k = 0.$$

Then, together with the relations between successive variables, this can all be expressed as the $kn \times kn$ linear eigenvalue problem

$$(3.1) \quad \underbrace{\left(\lambda \begin{bmatrix} A_k & 0 & \cdots & 0 \\ 0 & I_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_n \end{bmatrix} + \begin{bmatrix} A_{k-1} & A_{k-2} & \cdots & A_0 \\ -I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -I_n & 0 \end{bmatrix} \right)}_{= C_1(\lambda)} \begin{bmatrix} x_1 \\ \vdots \\ x_{k-1} \\ x_k \end{bmatrix} = 0.$$

Conversely, if we start with (3.1), then the last $k-1$ block rows immediately constrain any solution of (3.1) to have the form

$$\begin{bmatrix} x_1 \\ \vdots \\ x_{k-1} \\ x_k \end{bmatrix} = \begin{bmatrix} \lambda^{k-1}x \\ \vdots \\ \lambda x \\ x \end{bmatrix} = \Lambda \otimes x$$

for some vector $x \in \mathbb{F}^n$. Thus to solve (3.1) it is reasonable to restrict attention to products of the form $C_1(\lambda) \cdot (\Lambda \otimes x)$. However,

$$(3.2) \qquad C_1(\lambda) \cdot \left(\Lambda \otimes x \right) = \left[\left(P(\lambda)x \right)^T \quad 0 \quad \cdots \quad 0 \right]^T \quad \text{for all} \quad x \in \mathbb{F}^n,$$

and so any solution of (3.1) leads to a solution of the original problem $P(\lambda)x = 0$. Now observe that (3.2) is equivalent to the identity

$$(3.3) \qquad C_1(\lambda) \cdot \left(\Lambda \otimes I_n \right) = C_1(\lambda) \begin{bmatrix} \lambda^{k-1}I_n \\ \vdots \\ \lambda I_n \\ I_n \end{bmatrix} = \begin{bmatrix} P(\lambda) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = e_1 \otimes P(\lambda).$$

Thus to generalize the companion form we consider the set of all $kn \times kn$ matrix pencils $L(\lambda) = \lambda X + Y$ satisfying the property

$$(3.4) \qquad L(\lambda) \cdot (\Lambda \otimes I_n) = L(\lambda) \begin{bmatrix} \lambda^{k-1}I_n \\ \vdots \\ \lambda I_n \\ I_n \end{bmatrix} = \begin{bmatrix} v_1 P(\lambda) \\ \vdots \\ v_{k-1}P(\lambda) \\ v_k P(\lambda) \end{bmatrix} = v \otimes P(\lambda)$$

for some vector $v = [v_1, \ldots, v_k]^T \in \mathbb{F}^k$. This set of pencils will be denoted by $\mathbb{L}_1(P)$ as a reminder that it generalizes the first companion form of P. To work with property (3.4) more effectively we also introduce the notation

$$(3.5) \qquad\qquad\qquad \mathcal{V}_P = \{v \otimes P(\lambda) \; : \; v \in \mathbb{F}^k\}$$

for the set of all possible right-hand sides of (3.4). Thus we have the following definition.

DEFINITION 3.1. $\mathbb{L}_1(P) := \{L(\lambda) = \lambda X + Y : X, Y \in \mathbb{F}^{kn \times kn}, \; L(\lambda) \cdot (\Lambda \otimes I_n) \in \mathcal{V}_P\}$.

We will sometimes use the phrase "$L(\lambda)$ satisfies the right ansatz with vector v" or "v is the right ansatz vector for $L(\lambda)$" when $L(\lambda) \in \mathbb{L}_1(P)$ and the vector v in (3.4) is the focus of attention. We say "right" ansatz here because $L(\lambda)$ is multiplied on the right by the block column $\Lambda \otimes I_n$; later we introduce an analogous "left ansatz."

From the properties of the Kronecker product it is easy to see that $\mathcal{V}_P$ is a vector space isomorphic to $\mathbb{F}^k$, and consequently that $\mathbb{L}_1(P)$ is also a vector space.

PROPOSITION 3.2. *For any polynomial $P(\lambda)$, $\mathbb{L}_1(P)$ is a vector space over $\mathbb{F}$.*

Since $C_1(\lambda)$ is always in $\mathbb{L}_1(P)$, we see that $\mathbb{L}_1(P)$ is a nontrivial vector space for any matrix polynomial P.

Our next goal is to show that, like the companion forms, pencils in $\mathbb{L}_1(P)$ are easily constructible from the data in P. A consequence of this construction is a characterization of all the pencils in $\mathbb{L}_1(P)$ and a calculation of $\dim \mathbb{L}_1(P)$. To simplify the

discussion, we introduce the following new operation on block matrices as a convenient tool for working with products of the form $L(\lambda) \cdot (\Lambda \otimes I_n)$.

DEFINITION 3.3 (Column shifted sum). *Let X and Y be block matrices*

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & & \vdots \\ X_{k1} & \cdots & X_{kk} \end{bmatrix}, \qquad Y = \begin{bmatrix} Y_{11} & \cdots & Y_{1k} \\ \vdots & & \vdots \\ Y_{k1} & \cdots & Y_{kk} \end{bmatrix}$$

with blocks $X_{ij}, Y_{ij} \in \mathbb{F}^{n \times n}$. Then the column shifted sum *of X and Y is defined to be*

$$X \boxplus Y := \begin{bmatrix} X_{11} & \cdots & X_{1k} & 0 \\ \vdots & & \vdots & \vdots \\ X_{k1} & \cdots & X_{kk} & 0 \end{bmatrix} + \begin{bmatrix} 0 & Y_{11} & \cdots & Y_{1k} \\ \vdots & \vdots & & \vdots \\ 0 & Y_{k1} & \cdots & Y_{kk} \end{bmatrix},$$

where the zero blocks are also $n \times n$.

As an example, for the first companion form $C_1(\lambda) = \lambda X_1 + Y_1$ of $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$, the column shifted sum $X_1 \boxplus Y_1$ is just

$$\begin{bmatrix} A_k & 0 & \cdots & 0 \\ 0 & I_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_n \end{bmatrix} \boxplus \begin{bmatrix} A_{k-1} & A_{k-2} & \cdots & A_0 \\ -I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -I_n & 0 \end{bmatrix} = \begin{bmatrix} A_k & A_{k-1} & \cdots & A_0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Thus, the property $C_1(\lambda) \cdot (\Lambda \otimes I_n) = e_1 \otimes P(\lambda)$ from (3.3) translates in terms of the column shifted sum into $X_1 \boxplus Y_1 = e_1 \otimes [A_k \ A_{k-1} \ \cdots \ A_0]$. In fact, this shifted sum operation is specifically designed to imitate the product of a pencil $L(\lambda) = \lambda X + Y$ with the block column matrix $\Lambda \otimes I_n$, in the sense of the following lemma.

LEMMA 3.4. *Let $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ be an $n \times n$ matrix polynomial, and $L(\lambda) = \lambda X + Y$ a $kn \times kn$ pencil. Then for $v \in \mathbb{F}^k$,*

$$(3.6) \quad (\lambda X + Y) \cdot (\Lambda \otimes I_n) = v \otimes P(\lambda) \iff X \boxplus Y = v \otimes [A_k \ A_{k-1} \ \cdots \ A_0],$$

and so the space $\mathbb{L}_1(P)$ may be alternatively characterized as

$$(3.7) \qquad \mathbb{L}_1(P) = \{\lambda X + Y : X \boxplus Y = v \otimes [A_k \ A_{k-1} \ \cdots \ A_0], \ v \in \mathbb{F}^k\}.$$

The proof follows from a straightforward calculation which is omitted. The column shifted sum now allows us to directly construct all the pencils in $\mathbb{L}_1(P)$.

THEOREM 3.5 (Characterization of pencils in $\mathbb{L}_1(P)$). *Let $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ be an $n \times n$ matrix polynomial, and $v \in \mathbb{F}^k$ any vector. Then the set of pencils in $\mathbb{L}_1(P)$ with right ansatz vector v consists of all $L(\lambda) = \lambda X + Y$ such that*

$$X = \begin{bmatrix} \overset{n}{v \otimes A_k} & \overset{(k-1)n}{-W} \end{bmatrix} \quad and \quad Y = \begin{bmatrix} \overset{(k-1)n}{W + \left(v \otimes [\ A_{k-1} \ \cdots \ A_1]\right)} & \overset{n}{v \otimes A_0} \end{bmatrix},$$

with $W \in \mathbb{F}^{kn \times (k-1)n}$ chosen arbitrarily.

Proof. Consider the *multiplication map* $\mathcal{M}$ that is implicit in the definition of $\mathbb{L}_1(P)$:

$$\mathbb{L}_1(P) \xrightarrow{\mathcal{M}} \mathcal{V}_P,$$
$$L(\lambda) \longmapsto L(\lambda)\,(\Lambda \otimes I_n).$$

Clearly $\mathcal{M}$ is linear. To see that $\mathcal{M}$ is surjective, let $v \otimes P(\lambda)$ be an arbitrary element of $\mathcal{V}_P$ and construct

$$X_v = \begin{array}{c} \begin{array}{cc} \scriptstyle n & \scriptstyle (k-1)n \end{array} \\ \left[\begin{array}{cc} v \otimes A_k & 0 \end{array} \right] \end{array} \quad \text{and} \quad Y_v = \left[\begin{array}{cc} v \otimes \begin{array}{c} \scriptstyle (k-1)n \\ \left[\begin{array}{ccc} A_{k-1} & \cdots & A_1 \end{array} \right] \end{array} & \begin{array}{c} \scriptstyle n \\ v \otimes A_0 \end{array} \end{array} \right].$$

Then $X_v \boxplus Y_v = v \otimes [A_k \ A_{k-1} \ \cdots \ A_0]$, so by Lemma 3.4, $L_v(\lambda) := \lambda X_v + Y_v$ is an $\mathcal{M}$-preimage of $v \otimes P(\lambda)$. The set of all $\mathcal{M}$-preimages of $v \otimes P(\lambda)$ is then $L_v(\lambda) + \ker \mathcal{M}$, so all that remains is to compute $\ker \mathcal{M}$. By (3.6), the kernel of $\mathcal{M}$ consists of all pencils $\lambda X + Y$ satisfying $X \boxplus Y = 0$. The definition of the shifted sum then implies that X and Y must have the form

$$X = \begin{array}{c} \begin{array}{cc} \scriptstyle n & \scriptstyle (k-1)n \end{array} \\ \left[\begin{array}{cc} 0 & -W \end{array} \right] \end{array} \quad \text{and} \quad Y = \begin{array}{c} \begin{array}{cc} \scriptstyle (k-1)n & \scriptstyle n \end{array} \\ \left[\begin{array}{cc} W & 0 \end{array} \right] \end{array},$$

where $W \in \mathbb{F}^{kn \times (k-1)n}$ is arbitrary. This completes the proof. $\square$

COROLLARY 3.6. $\dim \mathbb{L}_1(P) = k(k-1)n^2 + k$.

Proof. Since $\mathcal{M}$ is surjective, $\dim \mathbb{L}_1(P) = \dim \ker \mathcal{M} + \dim \mathcal{V}_P = k(k-1)n^2 + k$. $\square$

Thus we see that $\mathbb{L}_1(P)$ is a relatively large subspace of the full pencil space (with dimension $2k^2n^2$), yet the pencils in $\mathbb{L}_1(P)$ are still easy to construct from the data in P. The next corollary isolates a special case of Theorem 3.5 that plays an important role in section 4.

COROLLARY 3.7. *Suppose $L(\lambda) = \lambda X + Y \in \mathbb{L}_1(P)$ has right ansatz vector $v = \alpha e_1$. Then*

$$(3.8) \qquad X = \left[\begin{array}{c|c} \alpha A_k & X_{12} \\ \hline 0 & -Z \end{array} \right] \quad \text{and} \quad Y = \left[\begin{array}{c|c} Y_{11} & \alpha A_0 \\ \hline Z & 0 \end{array} \right]$$

for some $Z \in \mathbb{F}^{(k-1)n \times (k-1)n}$.

Note that $C_1(\lambda)$ fits the pattern in Corollary 3.7 with $v = e_1$ and $Z = -I_{(k-1)n}$.

The second important property of the companion form is the simple relationship between its eigenvectors and those of the polynomial P that it linearizes. From the discussion following (3.1) it is evident that every eigenvector of $C_1(\lambda)$ has the form $\Lambda \otimes x$, where x is an eigenvector of P. Thus eigenvectors of P are recovered simply by extracting the last n coordinates from eigenvectors of the companion form. Our next result shows that linearizations in $\mathbb{L}_1(P)$ also have this property.

THEOREM 3.8 (Eigenvector Recovery Property for $\mathbb{L}_1(P)$). *Let $P(\lambda)$ be an $n \times n$ matrix polynomial of degree k, and $L(\lambda)$ any pencil in $\mathbb{L}_1(P)$ with nonzero right ansatz vector v. Then $x \in \mathbb{C}^n$ is an eigenvector for $P(\lambda)$ with finite eigenvalue $\lambda \in \mathbb{C}$ if and only if $\Lambda \otimes x$ is an eigenvector for $L(\lambda)$ with eigenvalue λ. If, in addition, P is regular and $L \in \mathbb{L}_1(P)$ is a linearization for P, then every eigenvector of L with finite eigenvalue λ is of the form $\Lambda \otimes x$ for some eigenvector x of P.*

Proof. The first statement follows immediately from the identity

$$L(\lambda)(\Lambda \otimes x) = L(\lambda)(\Lambda \otimes I_n)(1 \otimes x) = (v \otimes P(\lambda))(1 \otimes x) = v \otimes (P(\lambda)x).$$

For the second statement, assume that $\lambda \in \mathbb{C}$ is a finite eigenvalue of $L(\lambda)$ with geometric multiplicity m, and let $y \in \mathbb{C}^{kn}$ be any eigenvector of $L(\lambda)$ associated with λ. Since $L(\lambda)$ is a linearization of $P(\lambda)$, the geometric multiplicity of λ for $P(\lambda)$ is

also m. Let $x_1, \ldots, x_m$ be linearly independent eigenvectors of $P(\lambda)$ associated with λ, and define $y_i = \Lambda \otimes x_i$ for $i = 1, \ldots, m$. Then $y_1, \ldots, y_m$ are linearly independent eigenvectors for $L(\lambda)$ with eigenvalue λ, and so y must be a linear combination of $y_1, \ldots, y_m$. Thus y has the form $y = \Lambda \otimes x$ for some eigenvector $x \in \mathbb{C}^n$ for P. □

A result analogous to Theorem 3.8 is also valid for the eigenvalue ∞. Because additional arguments are needed, this will be deferred until section 4.

The above development and analysis of the pencil space $\mathbb{L}_1(P)$ has a parallel version in which the starting point is the second companion form $C_2(\lambda) = \lambda X_2 + Y_2$, as in (2.3). The analogue of (3.3) is the identity

$$\begin{bmatrix} \lambda^{k-1} I_n & \cdots & \lambda I_n & I_n \end{bmatrix} \cdot C_2(\lambda) = \begin{bmatrix} P(\lambda) & 0 & \cdots & 0 \end{bmatrix},$$

expressed more compactly as $(\Lambda^T \otimes I_n) \cdot C_2(\lambda) = e_1^T \otimes P(\lambda)$. This leads us to consider pencils $L(\lambda) = \lambda X + Y$ satisfying the "left ansatz"

$$(3.9) \qquad\qquad \left(\Lambda^T \otimes I_n\right) \cdot L(\lambda) = w^T \otimes P(\lambda),$$

and to a corresponding vector space $\mathbb{L}_2(P)$. The vector w in (3.9) will be referred to as the "left ansatz vector" for $L(\lambda)$.

DEFINITION 3.9. *With* $\mathcal{W}_P = \{w^T \otimes P(\lambda) : w \in \mathbb{F}^k\}$, *we define*

$$\mathbb{L}_2(P) = \{L(\lambda) = \lambda X + Y : X, Y \in \mathbb{F}^{kn \times kn}, \left(\Lambda^T \otimes I_n\right) \cdot L(\lambda) \in \mathcal{W}_P\}.$$

The analysis of $\mathbb{L}_2(P)$ is aided by the introduction of the following block matrix operation.

DEFINITION 3.10 (Row shifted sum). *Let* X *and* Y *be block matrices*

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & & \vdots \\ X_{k1} & \cdots & X_{kk} \end{bmatrix}, \qquad Y = \begin{bmatrix} Y_{11} & \cdots & Y_{1k} \\ \vdots & & \vdots \\ Y_{k1} & \cdots & Y_{kk} \end{bmatrix}$$

with blocks $X_{ij}, Y_{ij} \in \mathbb{F}^{n \times n}$. *Then the* row shifted sum *of* X *and* Y *is defined to be*

$$X \boxplus Y := \begin{bmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & & \vdots \\ X_{k1} & \cdots & X_{kk} \\ 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} 0 & \cdots & 0 \\ Y_{11} & \cdots & Y_{1k} \\ \vdots & & \vdots \\ Y_{k1} & \cdots & Y_{kk} \end{bmatrix},$$

where the zero blocks are also $n \times n$.

The following analogue of Lemma 3.4 establishes the correspondence between the left ansatz and row shifted sums.

LEMMA 3.11. *Let* $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ *be an* $n \times n$ *matrix polynomial, and* $L(\lambda) = \lambda X + Y$ *a* $kn \times kn$ *pencil. Then for any* $w \in \mathbb{F}^k$,

$$(3.10) \qquad (\Lambda^T \otimes I_n) \cdot (\lambda X + Y) = w^T \otimes P(\lambda) \iff X \boxplus Y = w^T \otimes \begin{bmatrix} A_k \\ \vdots \\ A_0 \end{bmatrix}.$$

Using these tools, one can characterize the pencils in $\mathbb{L}_2(P)$ in a manner completely analogous to Theorem 3.5, and thus conclude that

$$(3.11) \qquad\qquad \dim \mathbb{L}_2(P) = \dim \mathbb{L}_1(P) = k(k-1)n^2 + k.$$

It is also not difficult to establish a stronger relationship between the spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$, which again immediately implies (3.11). Here for a polynomial $P(\lambda) = \sum_{i=0}^k \lambda^i A_i$, we use P^T to denote the polynomial $\sum_{i=0}^k \lambda^i A_i^T$; by extension, if $\mathcal{S}$ is any set of polynomials, then $\mathcal{S}^T$ is the set $\{P^T : P \in \mathcal{S}\}$.

PROPOSITION 3.12. $\mathbb{L}_2(P) = [\mathbb{L}_1(P^T)]^T$.

Proof.

$$L \in \mathbb{L}_1(P^T) \Leftrightarrow L(\lambda) \cdot (\Lambda \otimes I_n) = v \otimes P^T(\lambda)$$
$$\Leftrightarrow \left(\Lambda^T \otimes I_n\right) \cdot L^T(\lambda) = v^T \otimes P(\lambda) \;\Leftrightarrow\; L^T \in \mathbb{L}_2(P). \qquad \square$$

The analogue of Theorem 3.8 for pencils in $\mathbb{L}_2(P)$ involves left eigenvectors of $P(\lambda)$ rather than right eigenvectors. Since the definition of a left eigenvector of a matrix polynomial does not seem to be completely standardized, we include here the definition used in this paper.

DEFINITION 3.13 (Left eigenvectors). *A left eigenvector of an $n \times n$ matrix polynomial P associated with a finite eigenvalue λ is a nonzero vector $y \in \mathbb{C}^n$ such that $y^* P(\lambda) = 0$. A left eigenvector for P corresponding to the eigenvalue ∞ is a left eigenvector for $\operatorname{rev} P$ associated with the eigenvalue 0.*

This definition differs from the one adopted in [5], although it is compatible with the usual definition for left eigenvectors of a matrix [6], [19]. We have chosen Definition 3.13 here because it is the one typically used in formulas for condition numbers of eigenvalues, a topic investigated in [8]. The following result shows that left eigenvectors of P are easily recovered from linearizations in $\mathbb{L}_2(P)$. The proof is completely analogous to that given for Theorem 3.8.

THEOREM 3.14 (Eigenvector Recovery Property for $\mathbb{L}_2(P)$). *Let $P(\lambda)$ be an $n \times n$ matrix polynomial of degree k, and $L(\lambda)$ any pencil in $\mathbb{L}_2(P)$ with nonzero left ansatz vector w. Then $y \in \mathbb{C}^n$ is a left eigenvector for $P(\lambda)$ with finite eigenvalue $\lambda \in \mathbb{C}$ if and only if $\overline{\Lambda} \otimes y$ is a left eigenvector for $L(\lambda)$ with eigenvalue λ. If, in addition, P is regular and $L \in \mathbb{L}_2(P)$ is a linearization for P, then every left eigenvector of L with finite eigenvalue λ is of the form $\overline{\Lambda} \otimes y$ for some left eigenvector y of P.*

Just as for Theorem 3.8, there is an analogous result for the eigenvalue ∞ that can be found in section 4.

In this section we have seen that pencils in $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ closely resemble the companion forms, and have eigenvectors that are simply related to those of P. Thus one can reasonably view $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ as large classes of "potential" linearizations for $P(\lambda)$. So far, though, we have not shown any of these "good candidates" to actually be linearizations. It is to this question that we turn next.

4. When is a pencil in $\mathbb{L}_1(P)$ a linearization? It is clear that not all pencils in the spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ are linearizations of P—consider, for example, any pencil in $\mathbb{L}_1(P)$ with right ansatz vector $v = 0$. In this section we focus on $\mathbb{L}_1(P)$ and obtain criteria for deciding whether a pencil from $\mathbb{L}_1(P)$ is a linearization for P or not. We show, for example, that for any given $L \in \mathbb{L}_1(P)$ there is typically a condition (specific to L) on the coefficient matrices of P that must be satisfied in order to guarantee that L is actually a linearization for P. Specific examples of such "linearization conditions" can be found in section 4.1 and in the tables in section 5. Analogues of all the results in this section also hold for $\mathbb{L}_2(P)$, with very similar arguments.

We begin with a result concerning the special case of the right ansatz (3.4) considered in Corollary 3.7. Note that P is not assumed here to be regular.

THEOREM 4.1. *Suppose that $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ with $A_k \neq 0$ is an $n \times n$ matrix polynomial, and $L(\lambda) = \lambda X + Y \in \mathbb{L}_1(P)$ has nonzero right ansatz vector $v = \alpha e_1$, so that*

$$(4.1) \qquad L(\lambda) \cdot (\Lambda \otimes I_n) = \alpha e_1 \otimes P(\lambda).$$

Partition X and Y as in (3.8) so that

$$(4.2) \qquad L(\lambda) \; = \; \lambda X + Y \; = \; \lambda \left[\begin{array}{c|c} \alpha A_k & X_{12} \\ \hline 0 & -Z \end{array}\right] + \left[\begin{array}{c|c} Y_{11} & \alpha A_0 \\ \hline Z & 0 \end{array}\right],$$

where $Z \in \mathbb{F}^{(k-1)n \times (k-1)n}$. Then Z nonsingular implies that $L(\lambda)$ is a strong linearization of $P(\lambda)$.

Proof. We show first that $L(\lambda)$ is a linearization of $P(\lambda)$. Begin the reduction of $L(\lambda)$ to $\mathrm{diag}(P(\lambda), I_{(k-1)n})$ using the unimodular matrix polynomials $T(\lambda)$ and $G(\lambda)$ defined in (2.5). In the product $L(\lambda)G(\lambda)$, clearly the first $k-1$ block-columns are the same as those of $L(\lambda)$; because the last block-column of $G(\lambda)$ is $\Lambda \otimes I$, we see from (4.1) that the last block-column of $L(\lambda)G(\lambda)$ is $\alpha e_1 \otimes P(\lambda)$. Partitioning Z in (4.2) into block columns $[Z_1\ Z_2\ \cdots\ Z_{k-1}]$, where $Z_i \in \mathbb{F}^{(k-1)n \times n}$, we thus obtain

$$L(\lambda)G(\lambda) \; = \; \begin{bmatrix} * & * & \cdots & * & * \\ Z_1 & (Z_2 - \lambda Z_1) & \cdots & (Z_{k-1} - \lambda Z_{k-2}) & -\lambda Z_{k-1} \end{bmatrix} G(\lambda)$$

$$= \; \begin{bmatrix} * & * & \cdots & * & \alpha P(\lambda) \\ Z_1 & (Z_2 - \lambda Z_1) & \cdots & (Z_{k-1} - \lambda Z_{k-2}) & 0 \end{bmatrix}.$$

Further transformation by block-column operations yields

$$L(\lambda)T(\lambda) = L(\lambda)\, \underbrace{G(\lambda) \begin{bmatrix} I & \lambda I & & \\ & I & & \\ & & I & \\ & & & \ddots \\ & & & & I \end{bmatrix} \begin{bmatrix} I & & & \\ & I & \lambda I & \\ & & I & \\ & & & \ddots \\ & & & & I \end{bmatrix} \cdots \begin{bmatrix} I & & & \\ & \ddots & & \\ & & I & \lambda I \\ & & & I \end{bmatrix}}_{=\, T(\lambda)} = \begin{bmatrix} * & \alpha P(\lambda) \\ \hline Z & 0 \end{bmatrix}.$$

Scaling and block-column permutations on $L(\lambda)T(\lambda)$ show that there exists a unimodular matrix polynomial $F(\lambda)$ such that

$$L(\lambda)F(\lambda) = \left[\begin{array}{c|c} P(\lambda) & W(\lambda) \\ \hline 0 & Z \end{array}\right]$$

for some matrix polynomial $W(\lambda)$. (Note that we have reached this point without any assumptions about Z.) Now if Z is nonsingular, then $L(\lambda)$ is a linearization for $P(\lambda)$, since

$$\left[\begin{array}{c|c} I & -W(\lambda)Z^{-1} \\ \hline 0 & Z^{-1} \end{array}\right] L(\lambda)F(\lambda) = \left[\begin{array}{c|c} P(\lambda) & 0 \\ \hline 0 & I_{(k-1)n} \end{array}\right].$$

To show that $L(\lambda)$ is also a strong linearization for $P(\lambda)$, it remains to show that $\mathrm{rev}\,L(\lambda) = \lambda Y + X$ is a linearization for $\mathrm{rev}\,P(\lambda)$. Now it would be nice if $\mathrm{rev}\,L(\lambda)$ were a pencil in $\mathbb{L}_1(\mathrm{rev}\,P)$, but it is not; however, a small modification of $\mathrm{rev}\,L(\lambda)$

is in $\mathbb{L}_1(\operatorname{rev} P)$. Observe that $\lambda^{k-1} \cdot \Lambda(1/\lambda) = [1, \lambda, \ldots, \lambda^{k-2}, \lambda^{k-1}]^T = R_k\Lambda$, where R_k denotes the $k \times k$ reverse identity matrix. Thus replacing λ by $1/\lambda$ in (4.1) and multiplying both sides by λ^k yields

$$\lambda L(1/\lambda) \cdot \left(\lambda^{k-1}\Lambda(1/\lambda) \otimes I\right) = \alpha e_1 \otimes \lambda^k P(1/\lambda),$$

or equivalently, $\operatorname{rev} L(\lambda) \cdot \left((R_k\Lambda) \otimes I\right) = \alpha e_1 \otimes \operatorname{rev} P(\lambda)$. Thus, $\widehat{L}(\lambda) := \operatorname{rev} L(\lambda) \cdot (R_k \otimes I)$ satisfies

$$(4.3) \qquad\qquad \widehat{L}(\lambda) \cdot (\Lambda \otimes I) = \alpha e_1 \otimes \operatorname{rev} P(\lambda),$$

and so $\widehat{L} \in \mathbb{L}_1(\operatorname{rev} P)$. (Observe that $\widehat{L}(\lambda)$ is just $\operatorname{rev} L(\lambda) = \lambda Y + X$ with the block-columns of Y and X arranged in reverse order.) Since $\widehat{L}$ and $\operatorname{rev} L$ are equivalent pencils, the proof will be complete once we show that $\lambda \widehat{X} + \widehat{Y} := \widehat{L}(\lambda)$ is a linearization for $\operatorname{rev} P(\lambda)$. However, $\widehat{X} = Y \cdot (R_k \otimes I)$ and $\widehat{Y} = X \cdot (R_k \otimes I)$, and hence from (4.2) it follows that

$$\widehat{X} = \left[\begin{array}{c|c} \alpha A_0 & \widehat{X}_{12} \\ \hline 0 & -\widehat{Z} \end{array}\right] \quad \text{and} \quad \widehat{Y} = \left[\begin{array}{c|c} \widehat{Y}_{11} & \alpha A_k \\ \hline \widehat{Z} & 0 \end{array}\right],$$

where $\widehat{Z} = -Z \cdot (R_{k-1} \otimes I)$. Clearly $\widehat{Z}$ is nonsingular if Z is, and so by the part of the theorem that has already been proved, $\widehat{L}$ (and therefore also $\operatorname{rev} L$) is a linearization for $\operatorname{rev} P(\lambda)$.   □

Remark 4.2. The fact (first proved in [4]) that the first companion form of any polynomial is always a strong linearization is a special case of Theorem 4.1.

When a matrix polynomial $P(\lambda)$ is regular, then it is easy to see from Definition 2.4 that any linearization for $P(\lambda)$ must also be regular. The next result shows something rather surprising: when a pencil L is in $\mathbb{L}_1(P)$ this minimal necessary condition of regularity is actually sufficient to guarantee that L is a linearization for P. This result serves to emphasize just how close a pencil is to being a linearization for P, even a strong linearization for P, once it satisfies the ansatz (3.4).

THEOREM 4.3 (Strong Linearization Theorem). *Let $P(\lambda)$ be a regular matrix polynomial, and let $L(\lambda) \in \mathbb{L}_1(P)$. Then the following statements are equivalent:*

(i) *$L(\lambda)$ is a linearization for $P(\lambda)$.*
(ii) *$L(\lambda)$ is a regular pencil.*
(iii) *$L(\lambda)$ is a strong linearization for $P(\lambda)$.*

Proof. "(i) $\Rightarrow$ (ii)": If $L(\lambda)$ is a linearization for $P(\lambda)$, then there exist unimodular matrix polynomials $E(\lambda)$, $F(\lambda)$ such that

$$E(\lambda)L(\lambda)F(\lambda) = \left[\begin{array}{cc} P(\lambda) & 0 \\ 0 & I_{(k-1)n} \end{array}\right].$$

Thus the regularity of $P(\lambda)$ implies the regularity of $L(\lambda)$.

"(ii) $\Rightarrow$ (iii)": Since $L(\lambda) \in \mathbb{L}_1(P)$, we know that $L(\lambda) \cdot (\Lambda \otimes I_n) = v \otimes P(\lambda)$ for some $v \in \mathbb{F}^k$. However, $L(\lambda)$ is regular, and so v is nonzero. Let $M \in \mathbb{F}^{k \times k}$ be any nonsingular matrix such that $Mv = \alpha e_1$. Then the regular pencil $\widetilde{L}(\lambda) := (M \otimes I_n) \cdot L(\lambda)$ is in $\mathbb{L}_1(P)$ with right ansatz vector αe_1, since

$$\begin{aligned} \widetilde{L}(\lambda)(\Lambda \otimes I_n) = (M \otimes I_n)L(\lambda)(\Lambda \otimes I_n) &= (M \otimes I_n)(v \otimes P(\lambda)) \\ &= Mv \otimes P(\lambda) \\ &= \alpha e_1 \otimes P(\lambda). \end{aligned}$$

Hence by Corollary 3.7 the matrices $\widetilde{X}$ and $\widetilde{Y}$ in $\widetilde{L}(\lambda) := \lambda\widetilde{X} + \widetilde{Y}$ have the forms

$$
\widetilde{X} = \begin{array}{c} \\ \left[\begin{array}{c|c} \overset{n}{\alpha A_k} & \overset{(k-1)n}{\widetilde{X}_{12}} \\ \hline 0 & -\widetilde{Z} \end{array}\right] \begin{array}{c} n \\ \\ (k-1)n \end{array} \end{array}
\quad \text{and} \quad
\widetilde{Y} = \begin{array}{c} \\ \left[\begin{array}{c|c} \overset{(k-1)n}{\widetilde{Y}_{11}} & \overset{n}{\alpha A_0} \\ \hline \widetilde{Z} & 0 \end{array}\right] \begin{array}{c} n \\ \\ (k-1)n \end{array} \end{array}.
$$

Now if $\widetilde{Z}$ was singular, there would exist a nonzero vector $w \in \mathbb{F}^{(k-1)n}$ such that $w^T\widetilde{Z} = 0$. But this would imply that

$$
\begin{bmatrix} 0 & w^T \end{bmatrix}(\lambda\widetilde{X} + \widetilde{Y}) = 0 \quad \text{for all} \ \lambda \in \mathbb{F},
$$

contradicting the regularity of $\widetilde{L}(\lambda)$. Thus $\widetilde{Z}$ is nonsingular, and so by Theorem 4.1 we know that $\widetilde{L}(\lambda)$, and hence also $L(\lambda)$, is a strong linearization for $P(\lambda)$.

"(iii) $\Rightarrow$ (i)" is trivial. $\square$

Now recall from Definitions 2.3 and 3.13 that a vector $x \in \mathbb{C}^n$ is a right (left) eigenvector for a polynomial P with eigenvalue ∞ if and only if x is a right (left) eigenvector for $\mathrm{rev}\,P$ with eigenvalue 0. Translating statements about infinite eigenvalues to ones about zero eigenvalues allows us to use Theorems 3.8, 3.14, and 4.3 to extend the eigenvector recovery properties of $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ to the eigenvalue ∞.

THEOREM 4.4 (Eigenvector Recovery at ∞). *Let $P(\lambda)$ be an $n \times n$ matrix polynomial of degree k, and $L(\lambda)$ any pencil in $\mathbb{L}_1(P)$ (resp., $\mathbb{L}_2(P)$) with nonzero right (left) ansatz vector v. Then $x \in \mathbb{C}^n$ is a right (left) eigenvector for $P(\lambda)$ with eigenvalue ∞ if and only if $e_1 \otimes x$ is a right (left) eigenvector for $L(\lambda)$ with eigenvalue ∞. If, in addition, P is regular and $L \in \mathbb{L}_1(P)$ (resp., $\mathbb{L}_2(P)$) is a linearization for P, then every right (left) eigenvector of L with eigenvalue ∞ is of the form $e_1 \otimes x$ for some right (left) eigenvector x of P with eigenvalue ∞.*

Proof. We give the proof only for right eigenvectors of $L \in \mathbb{L}_1(P)$ here. The argument for recovery of left eigenvectors of $L \in \mathbb{L}_2(P)$ is essentially the same, given the analogues of Theorems 4.1 and 4.3 for $\mathbb{L}_2(P)$.

For any $L(\lambda)$ define $\widehat{L}(\lambda) := \mathrm{rev}\,L(\lambda) \cdot (R_k \otimes I)$. Then the reasoning used in Theorem 4.1 to obtain (4.3) shows that $L \in \mathbb{L}_1(P) \Rightarrow \widehat{L} \in \mathbb{L}_1(\mathrm{rev}\,P)$, with the same nonzero right ansatz vector v. By Theorem 3.8 we know that x is a right eigenvector for $\mathrm{rev}\,P$ with eigenvalue 0 if and only if $\Lambda \otimes x = e_k \otimes x$ is a right eigenvector for $\widehat{L}$ with eigenvalue 0. However, $e_k \otimes x$ is a right eigenvector for $\widehat{L}$ if and only if $e_1 \otimes x = (R_k \otimes I)(e_k \otimes x)$ is a right eigenvector for $\mathrm{rev}\,L$, both with eigenvalue 0. This establishes the first part of the theorem.

If P is regular and $L \in \mathbb{L}_1(P)$ is a linearization for P, then by Theorem 4.3 $\widehat{L} \in \mathbb{L}_1(\mathrm{rev}\,P)$ is a linearization for $\mathrm{rev}\,P$. Theorem 3.8 then implies that every right eigenvector of $\widehat{L}$ with eigenvalue 0 is of the form $e_k \otimes x$, where x is a right eigenvector of $\mathrm{rev}\,P$ with eigenvalue 0; equivalently every right eigenvector of $\mathrm{rev}\,L$ with eigenvalue 0 is of the form $e_1 \otimes x$ for some right eigenvector x of $\mathrm{rev}\,P$ with eigenvalue 0. This establishes the second part of the theorem. $\square$

4.1. Linearization conditions. A useful by-product of the proof of Theorem 4.3 is a simple procedure for generating a symbolic "linearization condition" for any given pencil $L \in \mathbb{L}_1(P)$, i.e., a necessary and sufficient condition (in terms of the data in P) for L to be a linearization for P. We describe this procedure and then illustrate with some examples.

PROCEDURE TO DETERMINE THE LINEARIZATION CONDITION FOR A PENCIL IN $\mathbb{L}_1(P)$:

(1) Suppose that $P(\lambda)$ is a regular polynomial and $L(\lambda) = \lambda X + Y \in \mathbb{L}_1(P)$ has nonzero right ansatz vector $v \in \mathbb{F}^k$, i.e., $L(\lambda) \cdot (\Lambda \otimes I_n) = v \otimes P(\lambda)$.

(2) Select any nonsingular matrix M such that $Mv = \alpha e_1$.

(3) Apply the corresponding block-transformation $M \otimes I_n$ to $L(\lambda)$ to produce $\widetilde{L}(\lambda) := (M \otimes I_n)L(\lambda)$, which must be of the form

$$(4.4) \qquad \widetilde{L}(\lambda) = \lambda \widetilde{X} + \widetilde{Y} = \lambda \left[\begin{array}{c|c} \widetilde{X}_{11} & \widetilde{X}_{12} \\ \hline 0 & -Z \end{array} \right] + \left[\begin{array}{c|c} \widetilde{Y}_{11} & \widetilde{Y}_{12} \\ \hline Z & 0 \end{array} \right],$$

where $\widetilde{X}_{11}$ and $\widetilde{Y}_{12}$ are $n \times n$. Since only Z is of interest here, it suffices to form just $\widetilde{Y} = (M \otimes I_n)Y$.

(4) Extract $\boxed{\det Z \neq 0}$, the linearization condition for $L(\lambda)$.

Note that this procedure can readily be implemented as a numerical algorithm to check if a pencil in $\mathbb{L}_1(P)$ is a linearization: choose M to be unitary, e.g., a Householder reflector, then use a rank revealing factorization such as the QR-decomposition with column pivoting or the singular value decomposition to check if Z is nonsingular.

Example 4.5. Consider the general quadratic polynomial $P(\lambda) = \lambda^2 A + \lambda B + C$ (assumed to be regular) and the following pencils in $\mathbb{L}_1(P)$:

$$L_1(\lambda) = \lambda \begin{bmatrix} A & B+C \\ A & 2B-A \end{bmatrix} + \begin{bmatrix} -C & C \\ A-B & C \end{bmatrix}, \quad L_2(\lambda) = \lambda \begin{bmatrix} 0 & -B \\ A & B-C \end{bmatrix} + \begin{bmatrix} B & 0 \\ C & C \end{bmatrix}.$$

Since

$$\begin{bmatrix} A & B+C \\ A & 2B-A \end{bmatrix} \boxplus \begin{bmatrix} -C & C \\ A-B & C \end{bmatrix} = \begin{bmatrix} A & B & C \\ A & B & C \end{bmatrix},$$

we have $L_1(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector $v = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. Subtracting the first entry from the second reduces v to e_1, and the corresponding block-row-operation on Y yields

$$\widetilde{Y} = \begin{bmatrix} -C & C \\ A-B+C & 0 \end{bmatrix}.$$

Hence $Z = A - B + C$, and $\det(A - B + C) = \det P(-1) \neq 0$ is the linearization condition. Thus $L_1(\lambda)$ is a linearization for P if and only if $\lambda = -1$ is *not* an eigenvalue of P. On the other hand, for $L_2(\lambda)$ we have

$$\begin{bmatrix} 0 & -B \\ A & B-C \end{bmatrix} \boxplus \begin{bmatrix} B & 0 \\ C & C \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ A & B & C \end{bmatrix},$$

and so $L_2(\lambda) \in \mathbb{L}_1(P)$ with $v = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$. Permuting the entries of v gives e_1, and applying the analogous block-row-permutation to Y yields

$$\widetilde{Y} = \begin{bmatrix} C & C \\ B & 0 \end{bmatrix}.$$

Thus $Z = \widetilde{Y}_{21} = B$, and so $\det B \neq 0$ is the linearization condition for $L_2(\lambda)$.

The next example shows that the linearization condition for a pencil in $\mathbb{L}_1(P)$ may depend on some nonlinear combination of the data in P, and thus its meaning may not be so easy to interpret.

Example 4.6. Consider the general cubic polynomial $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$ (again assumed to be regular) and the pencil

$$L_3(\lambda) = \lambda X + Y = \lambda \begin{bmatrix} A & 0 & 2C \\ -2A & -B-C & D-4C \\ 0 & A & -I \end{bmatrix} + \begin{bmatrix} B & -C & D \\ C-B & 2C-D & -2D \\ -A & I & 0 \end{bmatrix}$$

in $\mathbb{L}_1(P)$. Since $X \boxplus Y = \begin{bmatrix} 1 & -2 & 0 \end{bmatrix}^T \otimes \begin{bmatrix} A & B & C & D \end{bmatrix}$, we have $v = \begin{bmatrix} 1 & -2 & 0 \end{bmatrix}^T$. Adding twice the first block-row of Y to the second block-row of Y gives

$$Z = \begin{bmatrix} B+C & -D \\ -A & I \end{bmatrix},$$

and hence the linearization condition $\det Z = \det(B + C - DA) \neq 0$. (Recall that for $n \times n$ blocks W, X, Y, Z with $YZ = ZY$, we have $\det \begin{bmatrix} W & X \\ Y & Z \end{bmatrix} = \det(WZ - XY)$. See [11].)

We have seen in this section that each pencil in $\mathbb{L}_1(P)$ has its own particular condition on the coefficient matrices of P that must be satisfied in order for the pencil to be a linearization for P. From this point of view it seems conceivable that there could be polynomials P for which very few of the pencils in $\mathbb{L}_1(P)$ are actually linearizations for P. However, the following result shows that this never happens; when P is regular the "bad" pencils in $\mathbb{L}_1(P)$ always form a very sparse subset of $\mathbb{L}_1(P)$.

THEOREM 4.7 (Linearizations Are Generic in $\mathbb{L}_1(P)$). *For any regular $n \times n$ matrix polynomial $P(\lambda)$ of degree k, almost every pencil in $\mathbb{L}_1(P)$ is a linearization for $P(\lambda)$. (Here by "almost every" we mean for all but a closed, nowhere dense set of measure zero in $\mathbb{L}_1(P)$.)*

Proof. Let $d = \dim \mathbb{L}_1(P) = k + (k-1)kn^2$, and let $L_1(\lambda), L_2(\lambda), \ldots, L_d(\lambda)$ be any fixed basis for $\mathbb{L}_1(P)$. Since any $L(\lambda) \in \mathbb{L}_1(P)$ can be uniquely expressed as a linear combination

$$L(\lambda) \;=\; \beta_1 L_1(\lambda) + \beta_2 L_2(\lambda) + \cdots + \beta_d L_d(\lambda),$$

we can view $\det L(\lambda)$ as a polynomial in λ whose coefficients $c_0, c_1, c_2, \ldots, c_{kn}$ are each polynomial functions of $\beta_1, \ldots, \beta_d$, that is, $c_i = c_i(\beta_1, \ldots, \beta_d)$.

Now by Theorem 4.3 we know that $L(\lambda) \in \mathbb{L}_1(P)$ *fails* to be a linearization for $P(\lambda)$ if and only if $\det L(\lambda) \equiv 0$, equivalently if all the coefficients c_i are zero. Thus the subset of pencils in $\mathbb{L}_1(P)$ that are not linearizations for $P(\lambda)$ can be characterized as the common zero set $\mathcal{Z}$ of the polynomials $\{c_i(\beta_1, \beta_2, \ldots, \beta_d) : 0 \leq i \leq kn\}$, i.e., as an algebraic subset of $\mathbb{F}^d$.

Since proper algebraic subsets of $\mathbb{F}^d$ are well known to be closed, nowhere dense subsets of measure zero, the proof will be complete once we show that $\mathcal{Z}$ is a proper subset of $\mathbb{F}^d$, or equivalently, that there is a pencil in $\mathbb{L}_1(P)$ that is a linearization for P. But this is immediate: the first companion form $C_1(\lambda)$ for $P(\lambda)$ is in $\mathbb{L}_1(P)$ and is always a linearization for P (see [5] or Remark 4.2). $\square$

Although $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ contain a large supply of linearizations for P, there do exist simple linearizations for P that are neither in $\mathbb{L}_1(P)$ nor in $\mathbb{L}_2(P)$. We illustrate this with a recent example from [1].

Example 4.8. For the cubic matrix polynomial $P(\lambda) = \lambda^3 A_3 + \lambda^2 A_2 + \lambda A_1 + A_0$, the pencil

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

is shown in [1] to be a linearization for P. Using shifted sums, it is easy to see that $L(\lambda)$ is in neither $\mathbb{L}_1(P)$ nor $\mathbb{L}_2(P)$.

4.2. Another view of $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. In section 3 we defined the pencil space $\mathbb{L}_1(P)$ by generalizing one particular property of the first companion form $C_1(\lambda)$ of P. A different connection between $\mathbb{L}_1(P)$ and $C_1(\lambda)$ can be established, which gives an alternative insight into why the pencils in $\mathbb{L}_1(P)$ retain so many of the attractive features of $C_1(\lambda)$. Using the first three steps of the procedure in section 4.1, together with the characterization of $\mathbb{L}_1(P)$-pencils given in Theorem 3.5 and Corollary 3.7, one can show that any $L(\lambda) \in \mathbb{L}_1(P)$ can be factored (non-uniquely) in the form

$$(4.5) \qquad L(\lambda) \;=\; (K \otimes I_n) \left[\begin{array}{c|c} \alpha I_n & U \\ \hline 0 & -Z \end{array} \right] C_1(\lambda),$$

where $Z \in \mathbb{F}^{(k-1)n \times (k-1)n}$ is the same as the block Z in Corollary 3.7 and (4.4), and $K \in \mathbb{F}^{k \times k}$ is nonsingular. Note that the scalar $\alpha \in \mathbb{F}$ is zero if and only if the right ansatz vector v of $L(\lambda)$ is zero. This factorization gives another reason why the *right* eigenvectors of pencils in $\mathbb{L}_1(P)$ have the same Kronecker product structure as those of $C_1(\lambda)$, and why pencils in $\mathbb{L}_1(P)$ are either strong linearizations of P (like $C_1(\lambda)$) or singular pencils, depending on the nonsingularity or singularity of the block Z and the scalar α.

In a completely analogous fashion one can factor any $L(\lambda) \in \mathbb{L}_2(P)$ as

$$(4.6) \qquad L(\lambda) \;=\; C_2(\lambda) \left[\begin{array}{c|c} \beta I_n & 0 \\ \hline T & -V \end{array} \right] (H \otimes I_n),$$

thus providing a different insight into the *left* eigenvector structure of pencils in $\mathbb{L}_2(P)$, and the fact that almost all pencils in $\mathbb{L}_2(P)$ are strong linearizations for P (like $C_2(\lambda)$).

On the other hand, certain aspects of $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ are less apparent from the point of view of these factorizations. For example, the fact that $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ are vector spaces is no longer so obvious. In addition, the criterion for a pencil to be an element of $\mathbb{L}_1(P)$ or $\mathbb{L}_2(P)$ is now implicit rather than explicit and is therefore rather harder to verify.

We are also interested in the possibility of the existence of pencils that are simultaneously in $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. The factored forms (4.5) and (4.6) might make it seem rather unlikely that there could be any nontrivial pencils in this intersection. However, in the next section we will see (using shifted sums) that this is an erroneous impression.

Finally, it is worth pointing out that the ansatz equations (3.4) and (3.9) enjoy the advantage of being identities in the variable λ, and so can be treated analytically as well as algebraically. This property is exploited in the analysis of the conditioning of eigenvalues of linearizations [8].

5. Double ansatz spaces. So far we have constructed two large vector spaces of pencils $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ for any given matrix polynomial $P(\lambda)$ and shown that when P is regular, almost all of these pencils are linearizations for P. Indeed, these spaces are so large that for any choice of (right or left) ansatz vector there are many degrees of freedom available for choosing a potential linearization in $\mathbb{L}_1(P)$ or $\mathbb{L}_2(P)$ with the given ansatz vector (see Theorem 3.5). This suggests that it might be possible to identify special subspaces of pencils in $\mathbb{L}_1(P)$ or $\mathbb{L}_2(P)$ with additional useful properties.

Recall that one of the key advantages of linearizations in $\mathbb{L}_1(P)$ is that right eigenvectors of P are easily recovered from right eigenvectors of the linearization. $\mathbb{L}_2(P)$ offers a similar advantage for recovery of left eigenvectors. Thus it seems natural to consider pencils in the intersection of $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$; for these pencils we can simply relate both the right and left eigenvectors of the pencil to those of the original polynomial P. This simultaneous eigenvector recovery property is particularly important in the investigation of the conditioning of linearizations [8]. Therefore we make the following definition.

DEFINITION 5.1 (Double ansatz spaces). *For any $n \times n$ matrix polynomial P of degree k, the double ansatz space of P is*

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

i.e., the set of $kn \times kn$ pencils $L(\lambda)$ that simultaneously satisfy

(5.1)   *a "right ansatz"* $L(\lambda) \cdot (\Lambda \otimes I) = v \otimes P(\lambda)$ *for some $v \in \mathbb{F}^k$,*

(5.2)   *and a "left ansatz"* $(\Lambda^T \otimes I) \cdot L(\lambda) = w^T \otimes P(\lambda)$ *for some $w \in \mathbb{F}^k$.*

The rest of this paper is devoted to developing some of the basic properties of $\mathbb{DL}(P)$-spaces, additional aspects of which are explored in [7], [8], [12]. In this section we characterize $\mathbb{DL}(P)$ and show how all the pencils in $\mathbb{DL}(P)$ may be constructed. In section 6 we reconsider the "linearization condition" discussed in section 4. As illustrated by Example 4.6, the intrinsic meaning of this condition can sometimes be rather obscure. However, we will see that for pencils in $\mathbb{DL}(P)$ this condition can always be expressed in a way that makes its meaning transparent.

A priori, the right and left ansatz vectors of a pencil in $\mathbb{DL}(P)$ may be any pair $v, w \in \mathbb{F}^k$. However, it turns out that only pairs with $v = w$ can ever be realized by a $\mathbb{DL}(P)$-pencil. To show this, we first need to determine when the equations $X \boxplus Y = S$ and $X \boxminus Y = T$ can be solved simultaneously for X and Y.

PROPOSITION 5.2. *Let $S = [S_{ij}]$ and $T = [T_{ji}]$ be block matrices of size $kn \times (k+1)n$ and $(k+1)n \times kn$, respectively, where $S_{ij}, T_{ji} \in \mathbb{F}^{n \times n}$ for $i = 1, \ldots, k$ and $j = 1, \ldots, k+1$. Then there exist block $k \times k$ matrices $X = [X_{ij}]$, $Y = [Y_{ij}]$ with blocks $X_{ij}, Y_{ij} \in \mathbb{F}^{n \times n}$ for $i, j = 1, \ldots, k$ such that*

(5.3) $X \boxplus Y = S$ *and* $X \boxminus Y = T$

if and only if for $j = 1, \ldots, k$ the blocks of S and T satisfy the compatibility conditions

(5.4) $$T_{jj} + \sum_{\mu=1}^{j-1}(T_{\mu,2j-\mu} - S_{\mu,2j-\mu}) = S_{jj} + \sum_{\mu=1}^{j-1}(S_{2j-\mu,\mu} - T_{2j-\mu,\mu})$$

and

(5.5) $$\sum_{\mu=1}^{j}(S_{\mu,2j+1-\mu} - T_{\mu,2j+1-\mu}) = \sum_{\mu=1}^{j}(T_{2j+1-\mu,\mu} - S_{2j+1-\mu,\mu}).$$

(*Here,* $S_{\nu,\eta} = 0 = T_{\eta,\nu}$ *whenever* $(\nu,\eta) \notin \{1,\ldots,k\} \times \{1,\ldots,k+1\}$.) *If* (5.3) *has a solution, then* X *and* Y *are uniquely determined by the formulas*

$$(5.6) \quad X_{ij} = T_{ij} + \sum_{\mu=1}^{i-1}(T_{\mu,j+i-\mu} - S_{\mu,j+i-\mu}), Y_{ij} = \sum_{\mu=1}^{i}(S_{\mu,j+i+1-\mu} - T_{\mu,j+i+1-\mu}),$$

$$(5.7) \quad X_{ji} = S_{ji} + \sum_{\mu=1}^{i-1}(S_{j+i-\mu,\mu} - T_{j+i-\mu,\mu}), Y_{ji} = \sum_{\mu=1}^{i}(T_{j+i+1-\mu,\mu} - S_{j+i+1-\mu,\mu}),$$

for $i,j = 1,\ldots,k$ *and* $j \geq i$.

Proof. Due to its technical nature, the proof is provided in Appendix A. □

We are now in a position to show not only that any $\mathbb{DL}(P)$-pencil has its right ansatz vector equal to its left ansatz vector, but also that every $v \in \mathbb{F}^k$ is actually realized as the ansatz vector of a pencil in $\mathbb{DL}(P)$, indeed of a unique pencil in $\mathbb{DL}(P)$. Note that this result does not require any regularity assumption on P.

THEOREM 5.3. *Let* $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ *be a matrix polynomial with coefficients in* $\mathbb{F}^{n \times n}$ *and* $A_k \neq 0$. *Then for vectors* $v = (v_1,\ldots,v_k)^T$ *and* $w = (w_1,\ldots,w_k)^T$ *in* $\mathbb{F}^k$ *there exists a* $kn \times kn$ *matrix pencil* $L(\lambda) = \lambda X + Y$ *that simultaneously satisfies*

$$(5.8) \quad L(\lambda) \cdot (\Lambda \otimes I) = v \otimes P(\lambda) \quad and \quad (\Lambda^T \otimes I) \cdot L(\lambda) = w^T \otimes P(\lambda)$$

if and only if $v = w$. *In this case, if* $X = [X_{ij}]$ *and* $Y = [Y_{ij}]$ *are written as block matrices with* $n \times n$ *blocks* X_{ij} *and* Y_{ij}, *then* X *and* Y *are uniquely determined by* v. *In particular, setting* $v_0 := 0$, $v_\mu := 0$, *and* $A_\mu := 0 \in \mathbb{F}^{n \times n}$ *for* $\mu < 0$ *or* $\mu > k$, *the blocks of* X *and* Y *satisfy the formulas*

$$(5.9) \quad X_{ij} = v_{\max(i,j)}A_{k+1-\min(i,j)} + \sum_{\mu=1}^{\min(i-1,j-1)}(v_{j+i-\mu}A_{k+1-\mu} - v_\mu A_{k+1-j-i+\mu}),$$

$$(5.10) \quad Y_{ij} = \sum_{\mu=1}^{\min(i,j)}(v_\mu A_{k-j-i+\mu} - v_{j+i+1-\mu}A_{k+1-\mu}), \qquad i,j = 1,\ldots,k.$$

Proof. See Appendix B for the proof. □

In light of the results in Theorem 5.3, we no longer need to refer separately to the right and left ansatz vectors of a pencil in $\mathbb{DL}(P)$. It suffices to say *the* ansatz vector v of $L \in \mathbb{DL}(P)$, and it is to be understood that v plays both roles.

We can also concisely summarize the result of Theorem 5.3 in a slightly different way. Viewing $\mathbb{DL}(P)$ as a special subspace of $\mathbb{L}_1(P)$, consider the multiplication map $\mathcal{M}$ (introduced in the proof of Theorem 3.5) restricted to the subspace $\mathbb{DL}(P)$. Then the following is an immediate corollary of Theorem 5.3.

COROLLARY 5.4. *For any polynomial* P, *the map* $\mathbb{DL}(P) \xrightarrow{\mathcal{M}} \mathcal{V}_P$ *is an isomorphism.*

Thus once an ansatz vector v has been chosen, a pencil from $\mathbb{DL}(P)$ is uniquely determined and can be computed using the formulas of Theorem 5.3.

Another significant property of $\mathbb{DL}(P)$ is worth mentioning here. A matrix polynomial is *symmetric* when all its coefficient matrices are symmetric. For symmetric P, a simple argument shows that every pencil in $\mathbb{DL}(P)$ is also symmetric: $L \in \mathbb{DL}(P)$ with ansatz vector v implies that L^T is also in $\mathbb{DL}(P)$ with the *same* ansatz vector v, and then $L = L^T$ follows from the uniqueness statement of Theorem 5.3.

TABLE 1

Some pencils in $\mathbb{DL}(P)$ for the general quadratic $P(\lambda) = \lambda^2 A + \lambda B + C$. Linearization condition found using procedure in section 4.1.

v	$L(\lambda) \in \mathbb{DL}(P)$ for given v	Linearization condition
$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\lambda \begin{bmatrix} A & 0 \\ 0 & -C \end{bmatrix} + \begin{bmatrix} B & C \\ C & 0 \end{bmatrix}$	$\det(C) \neq 0$
$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} 0 & A \\ A & B \end{bmatrix} + \begin{bmatrix} -A & 0 \\ 0 & C \end{bmatrix}$	$\det(A) \neq 0$
$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} A & A \\ A & B-C \end{bmatrix} + \begin{bmatrix} B-A & C \\ C & C \end{bmatrix}$	$\det(A - B + C) = \det[P(-1)] \neq 0$
$\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$	$\lambda \begin{bmatrix} \alpha A & \beta A \\ \beta A & \beta B - \alpha C \end{bmatrix} + \begin{bmatrix} \alpha B - \beta A & \alpha C \\ \alpha C & \beta C \end{bmatrix}$	$\det(\beta^2 A - \alpha\beta B + \alpha^2 C) \neq 0$; equivalently, $\det\left[P(-\frac{\beta}{\alpha})\right] \neq 0$

TABLE 2

Some pencils in $\mathbb{DL}(P)$ for the general cubic $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$. Linearization condition found using procedure in section 4.1.

v	$L(\lambda) \in \mathbb{DL}(P)$ for given v	Linearization condition
$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$	$\lambda \begin{bmatrix} A & 0 & 0 \\ 0 & -C & -D \\ 0 & -D & 0 \end{bmatrix} + \begin{bmatrix} B & C & D \\ C & D & 0 \\ D & 0 & 0 \end{bmatrix}$	$\det D \neq 0$
$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	$\lambda \begin{bmatrix} 0 & A & 0 \\ A & B & 0 \\ 0 & 0 & -D \end{bmatrix} + \begin{bmatrix} -A & 0 & 0 \\ 0 & C & D \\ 0 & D & 0 \end{bmatrix}$	$\det A \cdot \det D \neq 0$
$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} 0 & 0 & A \\ 0 & A & B \\ A & B & C \end{bmatrix} + \begin{bmatrix} 0 & -A & 0 \\ -A & -B & 0 \\ 0 & 0 & D \end{bmatrix}$	$\det A \neq 0$
$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$	$\lambda \begin{bmatrix} A & 0 & -A \\ 0 & -A-C & -B-D \\ -A & -B-D & -C \end{bmatrix} + \begin{bmatrix} B & A+C & D \\ A+C & B+D & 0 \\ D & 0 & -D \end{bmatrix}$	$\det \begin{bmatrix} A+C & B+D \\ B+D & A+C \end{bmatrix} \neq 0$
$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} A & A & A \\ A & A+B-C & B-D \\ A & B-D & C-D \end{bmatrix} + \begin{bmatrix} B-A & C-A & D \\ C-A & C+D-B & D \\ D & D & D \end{bmatrix}$	$\det \begin{bmatrix} C-B & A-B+D \\ A-B+D & A-C+D \end{bmatrix} \neq 0$

Examples of pencils in $\mathbb{DL}(P)$ for $k = 2$ and $k = 3$ may be found in Tables 1 and 2. Using shifted sums, one easily verifies that these examples are indeed in both $\mathbb{L}_1(P)$ *and* $\mathbb{L}_2(P)$, with the *same* right and left ansatz vector v. Note that if A, B, C, and D are symmetric, then so are all the pencils in these examples. Symmetric linearizations are studied in more detail in [7].

Perhaps the most striking property of the space $\mathbb{DL}(P)$ is that the linearization condition for each pencil in $\mathbb{DL}(P)$ can be directly linked to its ansatz vector v, as will be seen in the next section.

6. The eigenvalue exclusion theorem. We now establish a connection between the linearization condition of any pencil $L \in \mathbb{DL}(P)$ and the ansatz vector v that defines L. For example, consider the cubic polynomial $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$ and the pencil

$$L(\lambda) = \lambda \begin{bmatrix} A & 0 & -A \\ 0 & -A-C & -B-D \\ -A & -B-D & -C \end{bmatrix} + \begin{bmatrix} B & A+C & D \\ A+C & B+D & 0 \\ D & 0 & -D \end{bmatrix}$$

in $\mathbb{DL}(P)$ with ansatz vector $v = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^T$. Using the procedure in section 4.1, one easily finds that

(6.1) $$\det \begin{bmatrix} A+C & B+D \\ B+D & A+C \end{bmatrix} \neq 0$$

is the linearization condition for $L(\lambda)$. (See also Table 2.) Now it is not immediately clear what the meaning of this condition is, or even whether it has any intrinsic meaning at all. However, the identity

$$\begin{bmatrix} 0 & I \\ I & I \end{bmatrix} \begin{bmatrix} A+C & B+D \\ B+D & A+C \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix}$$

$$= \begin{bmatrix} -A+B-C+D & A+C \\ 0 & A+B+C+D \end{bmatrix} = \begin{bmatrix} P(-1) & A+C \\ 0 & P(+1) \end{bmatrix}$$

shows that condition (6.1) is equivalent to saying that neither -1 nor $+1$ is an eigenvalue of the matrix polynomial $P(\lambda)$. Thus in this example we can reinterpret the linearization condition from section 4.1 as an "eigenvalue exclusion" condition.

Why do these particular eigenvalues need to be excluded? And what role, if any, does the ansatz vector $v = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^T$ play here? Observe that if we interpret the components of v as the coefficients of a scalar polynomial, then we obtain $x^2 - 1$, whose roots are exactly the eigenvalues that have to be excluded in order to guarantee that $L(\lambda)$ is a linearization for $P(\lambda)$. The goal of this section is to show that this is not merely a coincidence, but rather an instance of a general phenomenon described by the "eigenvalue exclusion theorem."

The main technical result needed to prove this theorem is an explicit formula for the determinant of a pencil $L(\lambda)$ in $\mathbb{DL}(P)$. To aid in the development of this formula we first introduce some notation to be used throughout this section. As before, $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ is an $n \times n$ matrix polynomial with nonzero leading coefficient A_k. The pencil $L(\lambda) \in \mathbb{DL}(P)$ under consideration has ansatz vector $v = [v_1, v_2, \ldots, v_k]^T$, with an associated scalar polynomial defined as follows.

DEFINITION 6.1 (v-polynomial). *With a vector $v = [v_1, v_2, \ldots, v_k]^T \in \mathbb{F}^k$ associate the scalar polynomial*

$$\mathsf{p}(x\,;v) = v_1 x^{k-1} + v_2 x^{k-2} + \cdots + v_{k-1}x + v_k,$$

referred to as the "v-polynomial" of the vector v. We adopt the convention that $\mathsf{p}(x\,;v)$ has a root at ∞ whenever $v_1 = 0$.

We also need to introduce the notion of the "Horner shifts" of a polynomial.

DEFINITION 6.2 (Horner shifts). *For any polynomial $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ and $0 \leq \ell \leq n$, the "degree ℓ Horner shift of $p(x)$" is $p_\ell(x) := a_n x^\ell + a_{n-1} x^{\ell-1} + \cdots + a_{n-\ell+1}x + a_{n-\ell}$.*

Remark 6.3. The polynomials in Definition 6.2 satisfy the recurrence relation

$$p_0(x) = a_n,$$
$$p_{\ell+1}(x) = xp_\ell(x) + a_{n-\ell-1} \quad \text{for } 0 \le \ell \le n-1,$$
$$p_n(x) = p(x),$$

and are precisely the polynomials appearing in Horner's method for evaluating the polynomial $p(x)$.

We have seen in Theorem 5.3 that $L(\lambda) \in \mathbb{DL}(P)$ is uniquely determined by the vector v and the polynomial P, so it is not surprising that one can also specify the columns of $L(\lambda)$ in terms of this data. This is done in the next lemma, where a block-column–wise description of $L(\lambda)$ is given. In this description we make extensive use of the standard $k \times k$ nilpotent Jordan block N from (2.4) in the matrix $N \otimes I$, employed here as a block-shift operator.

LEMMA 6.4 (Block-column structure of pencils in $\mathbb{DL}(P)$). *Suppose that* $L(\lambda) = \lambda X + Y$ *is in* $\mathbb{DL}(P)$ *with ansatz vector* v. *Partition* X *and* Y *as*

$$X = \begin{bmatrix} X_1 & X_2 & \cdots & X_k \end{bmatrix} \quad and \quad Y = \begin{bmatrix} Y_1 & \cdots & Y_{k-1} & Y_k \end{bmatrix},$$

where $X_\ell, Y_\ell \in \mathbb{F}^{nk \times n}$, $\ell = 1, \ldots, k$. *Then with* $Y_0 := 0$, *the block-columns* Y_ℓ *satisfy the recurrence*

$$(6.2) \qquad Y_\ell = (N \otimes I)(Y_{\ell-1} - v \otimes A_{k-\ell+1}) + v_\ell \begin{bmatrix} A_{k-1} \\ \vdots \\ A_0 \end{bmatrix}, \quad 1 \le \ell \le k-1,$$

$$(6.3) \qquad Y_k = v \otimes A_0.$$

The block-columns of X *are determined by* $X_\ell = -Y_{\ell-1} + v \otimes A_{k-\ell+1}$ *for* $1 \le \ell \le k$, *and the pencil* $L(\lambda)$ *has the columnwise description*

$$(6.4) \;\; L(\lambda) = \begin{bmatrix} Y_1 & Y_2 - \lambda Y_1 & & Y_{k-1} - \lambda Y_{k-2} & v \otimes A_0 - \lambda Y_{k-1} \\ + \lambda v \otimes A_k & + \lambda v \otimes A_{k-1} & \cdots & + \lambda v \otimes A_2 & + \lambda v \otimes A_1 \end{bmatrix}.$$

Proof. Let $Y_0 = [Y_{i0}] := 0$, $X_\ell = [X_{i\ell}]$, and $Y_\ell = [Y_{i\ell}]$ for $n \times n$ blocks $Y_{i0}, X_{i\ell}, Y_{i\ell}$, where $i = 1, \ldots, k$. Then we obtain from (5.10) for $1 \le i < \ell \le k-1$ that

$$Y_{i\ell} = \sum_{\mu=1}^{i}(v_\mu A_{k-\ell-i+\mu} - v_{\ell+i+1-\mu}A_{k+1-\mu})$$
$$= \sum_{\mu=1}^{i+1}(v_\mu A_{k-\ell-i+\mu} - v_{\ell+i+1-\mu}A_{k+1-\mu}) - v_{i+1}A_{k+1-\ell} + v_\ell A_{k-i}$$
$$= Y_{i+1,\ell-1} - v_{i+1}A_{k+1-\ell} + v_\ell A_{k-i}.$$

Analogously, we obtain for $1 \le \ell \le i \le k-1$ that

$$Y_{i\ell} = \sum_{\mu=1}^{\ell}(v_\mu A_{k-\ell-i+\mu} - v_{\ell+i+1-\mu}A_{k+1-\mu})$$
$$= \sum_{\mu=1}^{\ell-1}(v_\mu A_{k-\ell-i+\mu} - v_{\ell+i+1-\mu}A_{k+1-\mu}) + v_\ell A_{k-i} - v_{i+1}A_{k+1-\ell}$$
$$= Y_{i+1,\ell-1} - v_{i+1}A_{k+1-\ell} + v_\ell A_{k-i}.$$

Since formula (5.10) also implies $Y_{k\ell} = v_\ell A_0$, we obtain

$$
Y_\ell = \begin{bmatrix} Y_{1\ell} \\ \vdots \\ Y_{k-1,\ell} \\ Y_{k\ell} \end{bmatrix} = \begin{bmatrix} Y_{2,\ell-1} \\ \vdots \\ Y_{k,\ell-1} \\ 0 \end{bmatrix} - \begin{bmatrix} v_2 A_{k-\ell+1} \\ \vdots \\ v_k A_{k-\ell+1} \\ 0 \end{bmatrix} + \begin{bmatrix} v_\ell A_k \\ \vdots \\ v_\ell A_1 \\ v_\ell A_0 \end{bmatrix}
$$

$$
= (N \otimes I)(Y_{\ell-1} - v \otimes A_{k-\ell+1}) + v_\ell \begin{bmatrix} A_{k-1} \\ \vdots \\ A_0 \end{bmatrix}
$$

for $\ell = 1, \ldots, k-1$. Noting that (3.6) implies $Y_k = v \otimes A_0$ and $X_\ell + Y_{\ell-1} = v \otimes A_{k-\ell+1}$ for $\ell = 1, \ldots, k$, we immediately obtain (6.4). □

Using (6.2), we can now develop a concise formula describing the action of the block-row $\Lambda^T(x) \otimes I$ on the block-column Y_ℓ, where x is a scalar variable taking values in $\mathbb{C}$ and $\Lambda^T(x) := \begin{bmatrix} x^{k-1} & x^{k-2} & \cdots & x & 1 \end{bmatrix}$. This formula will be used repeatedly and plays a central role in the proof of Theorem 6.6. (Note that $\Lambda^T(x)v$ is the same as the scalar v-polynomial $\mathsf{p}(x\,;v)$.)

LEMMA 6.5. *Suppose that $L(\lambda) \in \mathbb{DL}(P)$ with ansatz vector v, and $\mathsf{p}(x\,;v)$ is the v-polynomial of v. Let Y_ℓ denote the ℓth block column of Y in $L(\lambda) = \lambda X + Y$, where $1 \le \ell \le k-1$. Then*

$$
(6.5) \qquad (\Lambda^T(x) \otimes I) Y_\ell = \mathsf{p}_{\ell-1}(x\,;v)P(x) - x\,\mathsf{p}(x\,;v)P_{\ell-1}(x),
$$

where $\mathsf{p}_{\ell-1}(x\,;v)$ and $P_{\ell-1}(\lambda)$ are the degree $\ell-1$ Horner shifts of $\mathsf{p}(x\,;v)$ and $P(\lambda)$, respectively.

Proof. The proof will proceed by induction on ℓ. First note that for the $k \times k$ nilpotent Jordan block N, it is easy to check that $\Lambda^T(x)N = \begin{bmatrix} 0 & x^{k-1} & \cdots & x \end{bmatrix} = x\Lambda^T(x) - x^k e_1^T$.

$\boxed{\ell = 1}$: Using (6.2), we have

$$
(\Lambda^T(x) \otimes I) Y_1 = (\Lambda^T(x) \otimes I) \left(v_1 \begin{bmatrix} A_{k-1} \\ \vdots \\ A_0 \end{bmatrix} - (N \otimes I)(v \otimes A_k) \right).
$$

Simplifying this gives

$$
\begin{aligned}
(\Lambda^T(x) \otimes I) Y_1 &= v_1 \left(P(x) - x^k A_k\right) - \left(\Lambda^T(x)N \otimes I\right)(v \otimes A_k) \\
&= v_1 P(x) - v_1 x^k A_k - \left(\left(x\Lambda^T(x) - x^k e_1^T\right)v \otimes A_k\right) \\
&= \mathsf{p}_0(x\,;v)P(x) - v_1 x^k A_k - \left(x\Lambda^T(x)v\right)A_k + \left(x^k e_1^T v\right)A_k \\
&= \mathsf{p}_0(x\,;v)P(x) - v_1 x^k A_k - x\,\mathsf{p}(x\,;v)A_k + v_1 x^k A_k \\
&= \mathsf{p}_0(x\,;v)P(x) - x\,\mathsf{p}(x\,;v)P_0(x),
\end{aligned}
$$

which establishes (6.5) for $\ell = 1$. The induction hypothesis is now the following:

$$
(6.6) \qquad (\Lambda^T(x) \otimes I) Y_{\ell-1} = \mathsf{p}_{\ell-2}(x\,;v)P(x) - x\,\mathsf{p}(x\,;v)P_{\ell-2}(x).
$$

$\boxed{\ell - 1 \Rightarrow \ell}$: Starting again with (6.2), we have

$$\left(\Lambda^T(x) \otimes I\right) Y_\ell = \left(\Lambda^T(x) \otimes I\right) \left((N \otimes I)(Y_{\ell-1} - v \otimes A_{k-\ell+1}) + v_\ell \begin{bmatrix} A_{k-1} \\ \vdots \\ A_0 \end{bmatrix}\right)$$

$$= \left(\Lambda^T(x)N \otimes I\right)(Y_{\ell-1} - v \otimes A_{k-\ell+1}) + v_\ell \left(\Lambda^T(x) \otimes I\right) \begin{bmatrix} A_{k-1} \\ \vdots \\ A_0 \end{bmatrix}$$

$$= \left(\left(x\Lambda^T(x) - x^k e_1^T\right) \otimes I\right)(Y_{\ell-1} - v \otimes A_{k-\ell+1}) + v_\ell \left(P(x) - x^k A_k\right)$$

$$= x\left(\Lambda^T(x) \otimes I\right) Y_{\ell-1} - x^k \left(e_1^T \otimes I\right) Y_{\ell-1} - \left(x\Lambda^T(x)v\right) A_{k-\ell+1}$$
$$+ v_1 x^k A_{k-\ell+1} + v_\ell P(x) - v_\ell x^k A_k.$$

Note that $\left(e_1^T \otimes I\right)Y_{\ell-1}$ is the topmost block in $Y_{\ell-1}$ and is equal to $v_1 A_{k-\ell+1} - v_\ell A_k$, by (5.10). Finally, invoking the induction hypothesis (6.6) gives

$$\left(\Lambda^T(x) \otimes I\right) Y_\ell = x\,\mathsf{p}_{\ell-2}(x\,;v)P(x) - x^2\,\mathsf{p}(x\,;v)P_{\ell-2}(x) - v_1 x^k A_{k-\ell+1} + v_\ell\, x^k A_k$$
$$- x\,\mathsf{p}(x\,;v)A_{k-\ell+1} + v_1 x^k A_{k-\ell+1} + v_\ell P(x) - v_\ell\, x^k A_k$$
$$= \left(x\,\mathsf{p}_{\ell-2}(x\,;v) + v_\ell\right)P(x) - x\,\mathsf{p}(x\,;v)\left(xP_{\ell-2}(x) + A_{k-\ell+1}\right)$$
$$= \mathsf{p}_{\ell-1}(x\,;v)P(x) - x\,\mathsf{p}(x\,;v)P_{\ell-1}(x).$$

This completes the proof. $\square$

THEOREM 6.6 (Determinant formula for pencils in $\mathbb{DL}(P)$). *Suppose that $L(\lambda)$ is in $\mathbb{DL}(P)$ with nonzero ansatz vector $v = [v_1, v_2, \ldots, v_k]^T$. Assume that v has m leading zeroes with $0 \leq m \leq k - 1$, so that $v_1 = v_2 = \cdots = v_m = 0$, $v_{m+1} \neq 0$ is the first nonzero coefficient of $\mathsf{p}(x\,;v)$, and $\mathsf{p}(x\,;v)$ has $k - m - 1$ finite roots in $\mathbb{C}$, counted with multiplicities, denoted here by $r_1, r_2, \ldots, r_{k-m-1}$. Then we have*

(6.7)
$$\det L(\lambda) = \begin{cases} (-1)^{n \cdot \lfloor \frac{k}{2} \rfloor}(v_1)^{kn} \det\left(P(r_1)P(r_2) \cdots P(r_{k-1})\right) \det P(\lambda) & \text{if } m = 0, \\ (-1)^s (v_{m+1})^{kn}(\det A_k)^m \det\left(P(r_1) \cdots P(r_{k-m-1})\right) \det P(\lambda) & \text{if } m > 0, \end{cases}$$

where $s = n\left(m + \lfloor \frac{m}{2} \rfloor + \lfloor \frac{k-m}{2} \rfloor\right)$.

Proof. The proof proceeds in three parts.

Part 1. We first consider the case when $m = 0$ (i.e., $v_1 \neq 0$) and $\mathsf{p}(x\,;v)$ has $k - 1$ *distinct* finite roots. The strategy of the proof is to reduce $L(\lambda)$ by a sequence of equivalence transformations to a point where the determinant can just be read off.

We begin the reduction process by right-multiplying $L(\lambda)$ by the block-Toeplitz matrix $T(\lambda)$. Recall that $T(\lambda)$ and $G(\lambda)$ denote the unimodular matrix polynomials defined in (2.5), and are related to each other via the factorization in (2.6). Using (6.4) for the description of $L(\lambda)$, an argument very similar to the one used in the proof of Theorem 4.1 yields the block-column–wise description

$$L(\lambda)G(\lambda) = \begin{bmatrix} Y_1 \\ + \lambda v \otimes A_k \end{bmatrix} \begin{vmatrix} Y_2 - \lambda Y_1 \\ + \lambda v \otimes A_{k-1} \end{vmatrix} \cdots \begin{vmatrix} Y_{k-1} - \lambda Y_{k-2} \\ + \lambda v \otimes A_2 \end{vmatrix} v \otimes P(\lambda) \end{bmatrix},$$

and hence

(6.8)
$$L(\lambda)T(\lambda) = \begin{bmatrix} Y_1 \\ + \lambda v \otimes P_0(\lambda) \end{bmatrix} \begin{vmatrix} Y_2 \\ + \lambda v \otimes P_1(\lambda) \end{vmatrix} \cdots \begin{vmatrix} Y_{k-1} \\ + \lambda v \otimes P_{k-2}(\lambda) \end{vmatrix} v \otimes P(\lambda) \end{bmatrix}.$$

Next we left-multiply by a constant (nonsingular) "Vandermonde-like" matrix M, built block-row–wise from $\Lambda^T(x) := [\, x^{k-1}\ x^{k-2}\ \cdots\ x\ 1\,]$ evaluated at each of the roots of $\mathsf{p}(x\,;v)$,

$$(6.9) \qquad M := \begin{bmatrix} e_1^T \\ \Lambda^T(r_1) \\ \Lambda^T(r_2) \\ \vdots \\ \Lambda^T(r_{k-1}) \end{bmatrix} \otimes I = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ r_1^{k-1} & r_1^{k-2} & \cdots & r_1 & 1 \\ r_2^{k-1} & r_2^{k-2} & \cdots & r_2 & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ r_{k-1}^{k-1} & r_{k-1}^{k-2} & \cdots & r_{k-1} & 1 \end{bmatrix} \otimes I.$$

Using Lemma 6.5 and the fact that $\Lambda^T(r_j)v = \mathsf{p}(r_j\,;v)$, we obtain that

$$\left(\Lambda^T(r_j) \otimes I\right)\left(Y_\ell + \lambda v \otimes P_{\ell-1}(\lambda)\right)$$
$$= \mathsf{p}_{\ell-1}(r_j\,;v)P(r_j) - r_j\,\mathsf{p}(r_j\,;v)P_{\ell-1}(r_j) + \lambda\,\mathsf{p}(r_j\,;v)P_{\ell-1}(\lambda).$$

Since $r_1, \ldots, r_{k-1}$ are the roots of $\mathsf{p}(x\,;v)$, the product $ML(\lambda)T(\lambda)$ simplifies to

$$\left[\begin{array}{cccc|c} * & * & \cdots & * & v_1 P(\lambda) \\ \hline \mathsf{p}_0(r_1\,;v)P(r_1) & \mathsf{p}_1(r_1\,;v)P(r_1) & \cdots & \mathsf{p}_{k-2}(r_1\,;v)P(r_1) & 0 \\ \mathsf{p}_0(r_2\,;v)P(r_2) & \mathsf{p}_1(r_2\,;v)P(r_2) & \cdots & \mathsf{p}_{k-2}(r_2\,;v)P(r_2) & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathsf{p}_0(r_{k-1}\,;v)P(r_{k-1}) & \mathsf{p}_1(r_{k-1}\,;v)P(r_{k-1}) & \cdots & \mathsf{p}_{k-2}(r_{k-1}\,;v)P(r_{k-1}) & 0 \end{array}\right].$$

This matrix now factors into

$$\underbrace{\left[\begin{array}{c|ccc} I & & & \\ \hline & P(r_1) & & \\ & & \ddots & \\ & & & P(r_{k-1}) \end{array}\right]}_{=:\,W} \left[\begin{array}{ccc|c} * & \cdots & * & v_1 P(\lambda) \\ \hline \mathsf{p}_0(r_1\,;v)I & \cdots & \mathsf{p}_{k-2}(r_1\,;v)I & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \mathsf{p}_0(r_{k-1}\,;v)I & \cdots & \mathsf{p}_{k-2}(r_{k-1}\,;v)I & 0 \end{array}\right],$$

and after reversing the order of the block-columns using $R \otimes I$, we have

$$(6.10) \qquad ML(\lambda)T(\lambda)(R \otimes I) = W \left[\begin{array}{c|c} v_1 P(\lambda) & * \\ 0 & \\ \vdots & V \otimes I \\ 0 & \end{array}\right],$$

where

$$V = \begin{bmatrix} \mathsf{p}_{k-2}(r_1\,;v) & \cdots & \mathsf{p}_1(r_1\,;v) & \mathsf{p}_0(r_1\,;v) \\ \vdots & \vdots & \vdots & \vdots \\ \mathsf{p}_{k-2}(r_{k-1}\,;v) & \cdots & \mathsf{p}_1(r_{k-1}\,;v) & \mathsf{p}_0(r_{k-1}\,;v) \end{bmatrix}$$

$$= \begin{bmatrix} (v_1 r_1^{k-2} + \cdots + v_{k-2}r_1 + v_{k-1}) & \cdots & (v_1 r_1 + v_2) & v_1 \\ \vdots & \vdots & \vdots & \vdots \\ (v_1 r_{k-1}^{k-2} + \cdots + v_{k-2}r_{k-1} + v_{k-1}) & \cdots & (v_1 r_{k-1} + v_2) & v_1 \end{bmatrix}.$$

All that remains is to observe that V can be reduced by $(\det = +1)$ column operations to

$$(6.11) \qquad v_1 \cdot \begin{bmatrix} r_1^{k-2} & r_1^{k-3} & \cdots & r_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{k-1}^{k-2} & r_{k-1}^{k-3} & \cdots & r_{k-1} & 1 \end{bmatrix},$$

so $\det(V \otimes I) = v_1^{(k-1)n} \det M$. Taking determinants on both sides of (6.10) now gives

$$\det M \;\cdot\; \det L(\lambda) \cdot \det T(\lambda) \cdot \det(R \otimes I)$$
$$= \det\left(P(r_1)P(r_2)\cdots P(r_{k-1})\right) \cdot \det\left(v_1 P(\lambda)\right) \cdot \det(V \otimes I).$$

Since

$$(6.12) \qquad \det(R \otimes I) = \det(R_k \otimes I_n) = (\det R_k)^n (\det I_n)^k = (-1)^{n \cdot \lfloor \frac{k}{2} \rfloor}$$

and $\det T(\lambda) = +1$, this simplifies to the desired result

$$(6.13) \qquad \det L(\lambda) \;=\; (-1)^{n \cdot \lfloor \frac{k}{2} \rfloor} (v_1)^{kn} \det\left(P(r_1)P(r_2)\cdots P(r_{k-1})\right) \det P(\lambda).$$

This completes the argument for the case when $m = 0$ and the $k - 1$ roots of $\mathsf{p}(x\,;v)$ are all distinct.

Part 2. We now describe how to modify this argument to handle $m > 0$, i.e., the first nonzero coefficient of $\mathsf{p}(x\,;v)$ is v_{m+1}. We will continue to assume that the $k - m - 1$ finite roots of $\mathsf{p}(x\,;v)$ are all distinct.

We start out the same way as before, postmultiplying $L(\lambda)$ by $T(\lambda)$ to get (6.8). But then, instead of M in (6.9), we use all available finite roots of $\mathsf{p}(x\,;v)$ to define the following modified version of M:

$$(6.14) \quad \widehat{M} := \begin{bmatrix} e_1^T \\ \vdots \\ e_{m+1}^T \\ \Lambda^T(r_1) \\ \vdots \\ \Lambda^T(r_{k-m-1}) \end{bmatrix} \otimes I_n \;=\; \left[\begin{array}{ccccc} & I_{m+1} & & 0 & \\ \hline r_1^{k-1} & r_1^{k-2} & \cdots & r_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{k-m-1}^{k-1} & r_{k-m-1}^{k-2} & \cdots & r_{k-m-1} & 1 \end{array} \right] \otimes I_n.$$

Now simplify the product $\widehat{M}L(\lambda)T(\lambda)$ using Lemma 6.5 and $\Lambda^T(r_\ell)v = \mathsf{p}(r_\ell\,;v) = 0$ as before, as well as the fact that $v_1 = v_2 = \cdots = v_m = 0$, which implies that $\mathsf{p}_0(x\,;v), \mathsf{p}_1(x\,;v), \ldots, \mathsf{p}_{m-1}(x\,;v)$ are all zero polynomials. Then we obtain

$\widehat{M}L(\lambda)T(\lambda)$

$$
= \left[\begin{array}{ccc|c}
 & & & 0 \\
 & * & & \vdots \\
 & & & 0 \\
\hline
* & \cdots & * & v_{m+1}P(\lambda) \\
\hline
\mathsf{p}_0(r_1\,;v)P(r_1) & \cdots & \mathsf{p}_{k-2}(r_1\,;v)P(r_1) & 0 \\
\vdots & \vdots & \vdots & \vdots \\
\mathsf{p}_0(r_{k-m-1}\,;v)P(r_{k-m-1}) & \cdots & \mathsf{p}_{k-2}(r_{k-m-1}\,;v)P(r_{k-m-1}) & 0
\end{array}\right]
$$

$$
= \underbrace{\left[\begin{array}{c|cccc}
 & & & & 0 \\
B & & * & & \vdots \\
 & & & & 0 \\
\hline
* & * & \cdots & * & v_{m+1}P(\lambda) \\
\hline
 & \mathsf{p}_m(r_1\,;v)P(r_1) & \cdots & \mathsf{p}_{k-2}(r_1\,;v)P(r_1) & 0 \\
0 & \vdots & \vdots & \vdots & \vdots \\
 & \mathsf{p}_m(r_{k-m-1}\,;v)P(r_{k-m-1}) & \cdots & \mathsf{p}_{k-2}(r_{k-m-1}\,;v)P(r_{k-m-1}) & 0
\end{array}\right]}_{\substack{mn \qquad\qquad (k-m-1)n \qquad\qquad n}},
$$

where the $mn \times mn$ block B can also be seen to have some further structure. First note that because of the structure of $\widehat{M}$, the block B in $\widehat{M}L(\lambda)T(\lambda)$ is exactly the same as the corresponding block in $L(\lambda)T(\lambda)$ in (6.8), which is just the first mn rows of

$$
\left[\begin{array}{c|c|c|c}
Y_1 & Y_2 & \cdots & Y_m \\
{}+\lambda v \otimes P_0(\lambda) & {}+\lambda v \otimes P_1(\lambda) & & {}+\lambda v \otimes P_{m-1}(\lambda)
\end{array}\right].
$$

But because $v_1 = v_2 = \cdots = v_m = 0$, the terms $\lambda v \otimes P_i(\lambda)$ make no contribution to these first mn rows. So B is the same as the first mn rows of

$$
[Y_1|Y_2|\cdots|Y_m].
$$

Using the recurrence (6.2) from Lemma 6.4 with $1 \le \ell \le m$, we can now show that B is actually block anti-triangular. When $\ell = 1$ we have $Y_1 = -Nv \otimes A_k$. Since the first m entries of Nv are $[v_2, v_3, \ldots, v_{m+1}]^T = [0, 0, \ldots, v_{m+1}]^T$, we see that the first block-column of B is $[0, \ldots, 0, -v_{m+1}A_k^T]^T$. With $\ell = 2$ we have $Y_2 = (N \otimes I)Y_1 - Nv \otimes A_{k-1}$, whose first mn rows are

$$
\begin{bmatrix} 0 \\ \vdots \\ 0 \\ -v_{m+1}A_k \\ * \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ -v_{m+1}A_{k-1} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -v_{m+1}A_k \\ * \end{bmatrix}.
$$

By induction, we then see that the first mn rows of Y_ℓ for $1 \le \ell \le m$ look like

$$
\left[0, \ldots, 0, -v_{m+1}A_k^T, *, \ldots, *\right]^T,
$$

with $m - \ell$ leading blocks of zeroes. Thus B has the block anti-triangular form

$$B = -v_{m+1} \cdot \begin{bmatrix} 0 & \cdots & 0 & A_k \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & * \\ 0 & A_k & \cdot^{\cdot^{\cdot}} & \vdots \\ A_k & * & \cdots & * \end{bmatrix},$$

and so $\widehat{M}L(\lambda)T(\lambda)$ is equal to

$$\left[\begin{array}{ccc|cccc|ccc} 0 & & -v_{m+1}A_k & & & & & & & 0 \\ & \cdot^{\cdot^{\cdot}} & & & & * & & & & \vdots \\ -v_{m+1}A_k & & * & & & & & & & 0 \\ \hline * & & & * & \cdots & & * & & & v_{m+1}P(\lambda) \\ \hline & & & \mathsf{p}_m(r_1\,;v)P(r_1) & & \cdots & & \mathsf{p}_{k-2}(r_1\,;v)P(r_1) & & 0 \\ & 0 & & \vdots & & \vdots & & \vdots & & \vdots \\ & & & \mathsf{p}_m(r_{k-m-1}\,;v)P(r_{k-m-1}) & & \cdots & & \mathsf{p}_{k-2}(r_{k-m-1}\,;v)P(r_{k-m-1}) & & 0 \end{array} \right].$$

Performing some block-column permutations gives us

(6.15)

$$\widehat{M}L(\lambda)T(\lambda)\left((R_m \oplus R_{k-m}) \otimes I_n\right)$$

$$= \left[\begin{array}{ccc|c|cccc} -v_{m+1}A_k & & 0 & 0 & & & & \\ & \cdot^{\cdot^{\cdot}} & & \vdots & & & * & \\ * & & -v_{m+1}A_k & 0 & & & & \\ \hline * & & & v_{m+1}P(\lambda) & & * & \cdots & * \\ \hline & & & 0 & \mathsf{p}_{k-2}(r_1\,;v)P(r_1) & & \cdots & \mathsf{p}_m(r_1\,;v)P(r_1) \\ & 0 & & \vdots & \vdots & & \vdots & \vdots \\ & & & 0 & \mathsf{p}_{k-2}(r_{k-m-1}\,;v)P(r_{k-m-1}) & & \cdots & \mathsf{p}_m(r_{k-m-1}\,;v)P(r_{k-m-1}) \end{array} \right],$$

which after factoring becomes

(6.16)

$$\left[\begin{array}{c|c|c} (-v_{m+1}I_m) \otimes I_n & 0 & 0 \\ \hline 0 & v_{m+1}I_n & 0 \\ \hline 0 & 0 & \widehat{W} \end{array} \right] \left[\begin{array}{ccc|c|c} A_k & & 0 & & \\ & \cdot^{\cdot^{\cdot}} & & 0 & * \\ * & & A_k & & \\ \hline & 0 & & P(\lambda) & * \\ \hline & 0 & & 0 & \widehat{V} \otimes I_n \end{array} \right],$$

where $\widehat{W} = \mathrm{diag}(P(r_1), \ldots, P(r_{k-m-1}))$ and

$$\widehat{V} = \begin{bmatrix} \mathsf{p}_{k-2}(r_1\,;v) & \cdots & \mathsf{p}_m(r_1\,;v) \\ \vdots & \vdots & \vdots \\ \mathsf{p}_{k-2}(r_{k-m-1}\,;v) & \cdots & \mathsf{p}_m(r_{k-m-1}\,;v) \end{bmatrix}$$

$$= \begin{bmatrix} (v_{m+1}r_1^{k-m-2} + \cdots + v_{k-1}) & \cdots & (v_{m+1}r_1 + v_{m+2}) & v_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ (v_{m+1}r_{k-m-1}^{k-m-2} + \cdots + v_{k-1}) & \cdots & (v_{m+1}r_{k-m-1} + v_{m+2}) & v_{m+1} \end{bmatrix}.$$

Since $v_{m+1} \neq 0$, this $(k-m-1) \times (k-m-1)$ matrix $\widehat{V}$ can be reduced by $(\det = +1)$ column operations in a manner analogous to the reduction of V in (6.11), so we see that

$$(6.17) \qquad \det(\widehat{V} \otimes I_n) \,=\, (v_{m+1})^{(k-m-1)n} \det \widehat{M} \,.$$

Now taking determinants on both sides of (6.15) and using the factorization (6.16) gives

$$\det \widehat{M} \cdot \det L(\lambda) \cdot \det T(\lambda) \cdot \det(R_m \otimes I_n) \cdot \det(R_{k-m} \otimes I_n)$$
$$= \det\left(P(r_1)P(r_2)\cdots P(r_{k-m-1})\right) \cdot \det(-v_{m+1}A_k)^m \cdot \det\left(v_{m+1}P(\lambda)\right) \cdot \det(\widehat{V} \otimes I_n).$$

Canceling $\det \widehat{M}$ on both sides using (6.17), and using $\det T(\lambda) = +1$ together with the fact that $\det(R \otimes I)$ is its own inverse, we get

$$\det L(\lambda) = \det\left(P(r_1)P(r_2)\cdots P(r_{k-m-1})\right) \cdot (-1)^{mn} \cdot (v_{m+1})^{kn} \cdot (\det A_k)^m$$
$$\cdot \det P(\lambda) \cdot \det(R_m \otimes I_n) \cdot \det(R_{k-m} \otimes I_n).$$

Finally, substituting $\det(R_m \otimes I_n) = (-1)^{n \cdot \lfloor \frac{m}{2} \rfloor}$ and $\det(R_{k-m} \otimes I_n) = (-1)^{n \cdot \lfloor \frac{k-m}{2} \rfloor}$ from (6.12) yields the desired formula (6.7). Note that this is consistent with formula (6.13) derived for the $m = 0$ case, as long as we interpret the term $(\det A_k)^m$ to be equal to $+1$ whenever $m = 0$, regardless of whether $\det A_k$ is zero or nonzero.

Part 3. Now that we know that (6.7) holds for any $v \in \mathbb{F}^k$ such that the corresponding $\mathsf{p}(x\,;v)$ has *distinct* finite roots, we can leverage this result to the general case by a continuity argument. For every *fixed m* and fixed polynomial $P(\lambda)$, the formula on the right-hand side of (6.7) is clearly a continuous function of the leading coefficient v_{m+1} and the roots $r_1, r_2, \ldots, r_{k-m-1}$ of $\mathsf{p}(x\,;v)$, and is defined for all lists in the set $\mathcal{D} = \left\{(v_{m+1}, r_1, r_2, \ldots, r_{k-m-1}) : v_{m+1} \neq 0\right\}$, regardless of whether the numbers $r_1, r_2, \ldots, r_{k-m-1}$ are distinct or not.

The left-hand side of (6.7) can also be viewed as a function defined and continuous for all lists in $\mathcal{D}$. To see this, first observe that the map

$$(v_{m+1}, r_1, r_2, \ldots, r_{k-m-1}) \;\mapsto\; (v_{m+1}, v_{m+2}, \ldots, v_k)$$

taking the leading coefficient and roots of the polynomial $\mathsf{p}(x\,;v)$ to the coefficients of the same polynomial $\mathsf{p}(x\,;v)$ is defined and continuous on $\mathcal{D}$, as well as being surjective. Next note that because of the isomorphism in Corollary 5.4, the unique pencil $L(\lambda) \in \mathbb{DL}(P)$ corresponding to $v = (0, 0, \ldots, 0, v_{m+1}, \ldots, v_k)^T$ can be expressed as a linear combination

$$L(\lambda) = v_{m+1}L_{m+1}(\lambda) + \cdots + v_kL_k(\lambda)$$

of the *fixed* pencils $L_i(\lambda)$ corresponding to $v = e_i$. Thus $\det L(\lambda)$ is a continuous function of $(v_{m+1}, v_{m+2}, \ldots, v_k)$, and hence also of $(v_{m+1}, r_1, r_2, \ldots, r_{k-m-1})$.

In summary, the two sides of (6.7) are continuous functions defined on the same domain $\mathcal{D}$ and have been shown to be equal on a *dense* subset

$$\left\{(v_{m+1}, r_1, r_2, \ldots, r_{k-m-1}) : v_{m+1} \neq 0 \;\text{ and }\; r_1, r_2, \ldots, r_{k-m-1} \;\text{ are distinct}\right\}$$

of $\mathcal{D}$. Therefore by continuity the two sides of (6.7) must be equal on all of $\mathcal{D}$. Since this argument holds for each m with $0 \leq m \leq k - 1$, the desired result is established for all nonzero $v \in \mathbb{F}^k$. $\quad\square$

We now have all the ingredients needed to prove the main result of this section. Keep in mind our convention that the "roots of $\mathsf{p}(x\,;v)$" includes a root at ∞ whenever $v_1 = 0$.

THEOREM 6.7 (Eigenvalue Exclusion Theorem). *Suppose that $P(\lambda)$ is a regular matrix polynomial and $L(\lambda)$ is in $\mathbb{DL}(P)$ with nonzero ansatz vector v. Then $L(\lambda)$ is a linearization for $P(\lambda)$ if and only if no root of the v-polynomial $\mathsf{p}(x\,;v)$ is an eigenvalue of $P(\lambda)$. (Note that this statement includes ∞ as one of the possible roots of $\mathsf{p}(x\,;v)$ or possible eigenvalues of $P(\lambda)$.)*

Proof. By Theorem 4.3, $L(\lambda)$ is a linearization for $P(\lambda)$ if and only if $L(\lambda)$ is regular. However, from the determinant formula (6.7) it follows that $L(\lambda)$ is regular if and only if no root of $\mathsf{p}(x\,;v)$ is an eigenvalue of $P(\lambda)$. $\square$

Using Theorem 6.7, we can now show that almost every pencil in $\mathbb{DL}(P)$ is a linearization for P. Although the same property was proved in Theorem 4.7 for pencils in $\mathbb{L}_1(P)$, the result for $\mathbb{DL}(P)$ is not a consequence of Theorem 4.7, since $\mathbb{DL}(P)$ is itself a closed, nowhere dense subset of measure zero in $\mathbb{L}_1(P)$. Neither can the proof of Theorem 4.7 be directly generalized in any simple way; hence the need for a different argument in the following result.

THEOREM 6.8 (Linearizations Are Generic in $\mathbb{DL}(P)$). *For any regular matrix polynomial $P(\lambda)$, pencils in $\mathbb{DL}(P)$ are linearizations of $P(\lambda)$ for almost all $v \in \mathbb{F}^k$. (Here "almost all" means for all but a closed, nowhere dense set of measure zero in $\mathbb{F}^k$.)*

Proof. Recall that the resultant [22] $\mathrm{res}(f, g)$ of two polynomials $f(x)$ and $g(x)$ is a polynomial in the coefficients of f and g with the property that $\mathrm{res}(f, g) = 0$ if and only if $f(x)$ and $g(x)$ have a common (finite) root. Now consider $\mathrm{res}(\mathsf{p}(x\,;v), \det P(x))$, which, because $P(\lambda)$ is fixed, can be viewed as a polynomial $r(v_1, v_2, \ldots, v_k)$ in the components of $v \in \mathbb{F}^k$. The zero set $\mathcal{Z}(r) = \left\{ v \in \mathbb{F}^k : r(v_1, v_2, \ldots, v_k) = 0 \right\}$, then, is exactly the set of $v \in \mathbb{F}^k$ for which some finite root of $\mathsf{p}(x\,;v)$ is an eigenvalue of $P(\lambda)$, together with the point $v = 0$. Recall that by our convention the v-polynomial $\mathsf{p}(x\,;v)$ has ∞ as a root exactly for $v \in \mathbb{F}^k$ lying in the hyperplane $v_1 = 0$. Thus by Theorem 6.7 the set of vectors $v \in \mathbb{F}^k$ for which the corresponding pencil $L(\lambda) \in \mathbb{DL}(P)$ is *not* a linearization of $P(\lambda)$ is either the proper algebraic set $\mathcal{Z}(r)$ or the union of two proper algebraic sets, $\mathcal{Z}(r)$ and the hyperplane $v_1 = 0$. However, the union of any finite number of proper algebraic sets is always a closed, nowhere dense set of measure zero in $\mathbb{F}^k$. $\square$

How far can the eigenvalue exclusion theorem be extended from $\mathbb{DL}(P)$-pencils to other pencils in $\mathbb{L}_1(P)$? Let us say that a pencil $L \in \mathbb{L}_1(P)$ with right ansatz vector v has the *eigenvalue exclusion property* if the statement "no root of the v-polynomial $\mathsf{p}(x\,;v)$ is an eigenvalue of $P(\lambda)$" is equivalent to the linearization condition for L. That there are pencils in $\mathbb{L}_1(P)$ with the eigenvalue exclusion property that are not in $\mathbb{DL}(P)$ is shown by the pencil $L_1(\lambda)$ in Example 4.5. The following variation of Example 4.6, though, is easily shown not to have the eigenvalue exclusion property.

Example 6.9. For the general cubic polynomial $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$ consider the pencil

$$L(\lambda) = \lambda X + Y = \lambda \begin{bmatrix} A & 0 & 2C \\ -2A & -B-C & A-4C \\ 0 & A & 0 \end{bmatrix} + \begin{bmatrix} B & -C & D \\ C-B & 2C-A & -2D \\ -A & 0 & 0 \end{bmatrix}$$

that is in $\mathbb{L}_1(P)$ but not in $\mathbb{DL}(P)$. Since $X \boxplus Y = \begin{bmatrix} 1 & -2 & 0 \end{bmatrix}^T \otimes \begin{bmatrix} A & B & C & D \end{bmatrix}$, the right ansatz vector is $v = \begin{bmatrix} 1 & -2 & 0 \end{bmatrix}^T$ with v-polynomial $\mathsf{p}(x\,;v) = x^2 - 2x$ and

roots 0 and 2. On the other hand, applying the procedure described in section 4.1 gives

$$Z = \begin{bmatrix} B + C & -A \\ -A & 0 \end{bmatrix},$$

and hence the linearization condition $\det Z = \det(-A^2) \neq 0$, equivalently $\det A \neq 0$. Thus $L(\lambda)$ is a linearization for $P(\lambda)$ if and only if ∞ is not an eigenvalue of $P(\lambda)$. In this example, then, the roots of the v-polynomial do not correctly predict the linearization condition for L.

The first companion form of a polynomial P is another example where the eigenvalue exclusion property is easily seen not to hold. Characterizing the set of pencils in $\mathbb{L}_1(P)$ for which the eigenvalue exclusion property does hold is an open problem.

7. Concluding remarks. By generalizing the first and second companion form linearizations for a matrix polynomial $P(\lambda)$, we have introduced two large vector spaces of pencils, $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$, which serve as sources of potential linearizations for $P(\lambda)$. The mild hypothesis that $P(\lambda)$ is regular makes almost every pencil in these spaces a linearization for $P(\lambda)$.

A number of properties enjoyed by the companion forms extend to the linearizations in $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$: they are strong linearizations, are readily constructed from the coefficient matrices of $P(\lambda)$, and have eigenvectors that reveal those of $P(\lambda)$. Furthermore, a simple procedure can be used to test when a pencil in $\mathbb{L}_1(P)$ or $\mathbb{L}_2(P)$ is a linearization of $P(\lambda)$.

The intersection of $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$, denoted by $\mathbb{DL}(P)$, is of particular significance. Pencils in $\mathbb{L}_1(P)$ reveal only right eigenvectors of $P(\lambda)$, while those in $\mathbb{L}_2(P)$ lead to left eigenvectors of $P(\lambda)$. Pencils in $\mathbb{DL}(P)$ therefore simultaneously reveal right as well as left eigenvectors of P. An isomorphism between $\mathbb{DL}(P)$ and $\mathbb{F}^k$ allows the association of a unique scalar polynomial of degree $k - 1$ to each pencil in $\mathbb{DL}(P)$. Linearizations in $\mathbb{DL}(P)$ can then be characterized by an eigenvalue exclusion property—a pencil in this distinguished subspace is a linearization precisely when no root of its associated scalar polynomial is an eigenvalue of P.

As remarked earlier, the first and second companion form linearizations have a significant drawback—they usually do not reflect any structure that may be present in $P(\lambda)$. Different linearizations can also exhibit very different conditioning. By systematizing the construction of large classes of linearizations that generalize the companion forms, we have provided a rich arena in which linearizations with additional properties like structure preservation or improved conditioning can be found. This is the subject of the work in [7], [8], [12].

Appendix A. Proof of Proposition 5.2.
Proof. "$\Rightarrow$": Assume that (5.3) holds. First, we show by induction on k that the formulas (5.6)–(5.7) hold.

$\boxed{k = 1}$: In this case, we have

$$X \boxrightarrow{} Y = S = \begin{bmatrix} S_{11} & S_{12} \end{bmatrix}, \qquad X \boxdownarrow{} Y = T = \begin{bmatrix} T_{11} \\ T_{21} \end{bmatrix}$$

and hence $X = S_{11} = T_{11}$ and $Y = S_{12} = T_{21}$, which coincides with (5.6)–(5.7).

$\boxed{k - 1 \Rightarrow k}$: By the definition of the column and row shifted sums, (5.3) implies

(A.1) $Y_{ik} = S_{i,k+1}$ and $Y_{ki} = T_{k+1,i}$

as well as $X_{ji} + Y_{j,i-1} = S_{ji}$ and $X_{ij} + Y_{i-1,j} = T_{ij}$ for $j = 1, \ldots, k$ and $i = 2, \ldots, k$, which together with (A.1) gives

$$(A.2) \qquad X_{ki} = S_{ki} - T_{k+1,i-1} \quad \text{and} \quad X_{ik} = T_{ik} - S_{i-1,k+1}$$

for $i = 1, \ldots, k$. (Remember that $S_{0,k+1} = 0 = T_{k+1,0}$ by convention.) In order to be able to use the induction hypothesis, let us partition X and Y as

$$X = \left[\begin{array}{ccc|c} & & & X_{1k} \\ & \widetilde{X} & & \vdots \\ & & & X_{k-1,k} \\ \hline X_{k1} & \ldots & X_{k,k-1} & X_{kk} \end{array}\right], \quad Y = \left[\begin{array}{ccc|c} & & & Y_{1k} \\ & \widetilde{Y} & & \vdots \\ & & & Y_{k-1,k} \\ \hline Y_{k1} & \ldots & Y_{k,k-1} & Y_{kk} \end{array}\right],$$

with $(n-1)k \times (n-1)k$ matrices $\widetilde{X}$ and $\widetilde{Y}$. Then we obtain

$$(A.3) \quad \widetilde{X} \boxplus\!\!\!\rightarrow \widetilde{Y} = \left[\begin{array}{cccc} S_{11} & \ldots & S_{1,k-1} & S_{1k} - X_{1k} \\ S_{21} & \ldots & S_{2,k-1} & S_{2k} - X_{2k} \\ \vdots & \ddots & \vdots & \vdots \\ S_{k-1,1} & \ldots & S_{k-1,k-1} & S_{k-1,k} - X_{k-1,k} \end{array}\right]$$

$$(A.4) \qquad = \left[\begin{array}{cccc} S_{11} & \ldots & S_{1,k-1} & S_{1k} - T_{1k} \\ S_{21} & \ldots & S_{2,k-1} & S_{2k} - T_{2k} + S_{1,k+1} \\ \vdots & \ddots & \vdots & \vdots \\ S_{k-1,1} & \ldots & S_{k-1,k-1} & S_{k-1,k} - T_{k-1,k} + S_{k-2,k+1} \end{array}\right] =: \widetilde{S}.$$

Analogously,

$$(A.5)$$

$$\widetilde{X} \boxplus\!\!\downarrow \widetilde{Y} = \underbrace{\left[\begin{array}{cccc} T_{11} & T_{12} & \ldots & T_{1,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ T_{k-1,1} & T_{k-1,2} & \ldots & T_{k-1,k-1} \\ T_{k1} - S_{k1} & T_{k2} - S_{k2} + T_{k+1,1} & \ldots & T_{k,k-1} - S_{k,k-1} + T_{k+1,k-2} \end{array}\right]}_{=: \widetilde{T}}.$$

Writing $\widetilde{S} = [\widetilde{S}_{ij}]$ and $\widetilde{T} = [\widetilde{T}_{ij}]$ with $n \times n$ blocks $\widetilde{S}_{ij}, \widetilde{T}_{ij}$ and using the induction hypothesis for $\widetilde{X} = [X_{ij}]$ and $\widetilde{Y} = [Y_{ij}]$, we then obtain for $i, j = 1, \ldots, k-1$ and $j \geq i$ that

$$(A.6) \quad X_{ij} = \widetilde{T}_{ij} + \sum_{\mu=1}^{i-1}(\widetilde{T}_{\mu,j+i-\mu} - \widetilde{S}_{\mu,j+i-\mu}), \quad Y_{ij} = \sum_{\mu=1}^{i}(\widetilde{S}_{\mu,j+i+1-\mu} - \widetilde{T}_{\mu,j+i+1-\mu}),$$

$$(A.7) \quad X_{ji} = \widetilde{S}_{ji} + \sum_{\mu=1}^{i-1}(\widetilde{S}_{j+i-\mu,\mu} - \widetilde{T}_{j+i-\mu,\mu}), \quad Y_{ji} = \sum_{\mu=1}^{i}(\widetilde{T}_{j+i+1-\mu,\mu} - \widetilde{S}_{j+i+1-\mu,\mu}),$$

where $\widetilde{S}_{\nu\eta} = 0 = \widetilde{T}_{\eta\nu}$ whenever $(\nu, \eta) \notin \{1, \ldots, k-1\} \times \{1, \ldots, k\}$. We claim that together with (A.1) and (A.2), the formulas (A.6)–(A.7) coincide with the formulas (5.6)–(5.7). We show this in detail for the first formula in (5.6); for the other

formulas there is a similar proof that is omitted. If $j + i \leq k$, then the block forms of $\widetilde{S}$ and $\widetilde{T}$ given in (A.4) and (A.5) immediately yield

$$X_{ij} = \widetilde{T}_{ij} + \sum_{\mu=1}^{i-1}(\widetilde{T}_{\mu,j+i-\mu} - \widetilde{S}_{\mu,j+i-\mu}) = T_{ij} + \sum_{\mu=1}^{i-1}(T_{\mu,j+i-\mu} - S_{\mu,j+i-\mu}).$$

If $j + i > k$ and $i, j < k$, then $j + i - m = k$ for some $m \geq 1$; using $S_{\nu\eta} = 0 = T_{\eta\nu}$ for $(\nu, \eta) \notin \{1, \ldots, k\} \times \{1, \ldots, k+1\}$, we obtain

$$X_{ij} = \widetilde{T}_{ij} + \sum_{\mu=1}^{i-1}(\widetilde{T}_{\mu,j+i-\mu} - \widetilde{S}_{\mu,j+i-\mu})$$

$$= \widetilde{T}_{ij} + \sum_{\mu=m+1}^{i-1}(\widetilde{T}_{\mu,j+i-\mu} - \widetilde{S}_{\mu,j+i-\mu}) - \widetilde{S}_{mk}$$

$$= T_{ij} + \sum_{\mu=m+1}^{i-1}(T_{\mu,j+i-\mu} - S_{\mu,j+i-\mu}) - S_{mk} + T_{mk} - S_{m-1,k+1}$$

$$= T_{ij} + \sum_{\mu=1}^{i-1}(T_{\mu,j+i-\mu} - S_{\mu,j+i-\mu}).$$

Finally, for $i = k$ or $j = k$ the statement follows immediately from (A.1) or (A.2). This concludes the inductive proof of the formulas (5.6)–(5.7). In particular, this implies that X and Y are uniquely determined by S and T. Note that X_{ii} and Y_{ii} now satisfy two distinct formulas for $i = 1, \ldots, n$. Since both right-hand sides in the formulas (5.6)–(5.7) must be equal in this case, we directly obtain (5.4) and (5.5).

"$\Leftarrow$": We have to show the existence of block-matrices $X = [X_{ij}]$ and $Y = [Y_{ij}]$ such that $X \boxplus\!\!\!\rightarrow Y = S$ and $X \boxplus\!\!\!\downarrow Y = T$. Define X_{ij} and Y_{ij} by the formulas (5.6)–(5.7). Because of (5.4) and (5.5), X and Y are well defined. We will now show in detail that $X \boxplus\!\!\!\rightarrow Y = S$. (The proof of $X \boxplus\!\!\!\downarrow Y = T$ is similar and will be omitted.) Indeed, formulas (5.6)–(5.7) imply $X_{j1} = S_{j1}$ and $Y_{ik} = S_{ik}$ for $i, j = 1, \ldots, k$. Moreover, we obtain for $i = 1, \ldots, k$ and $j = 2, \ldots, k$ that

$$X_{ij} + Y_{i,j-1} = T_{ij} + \sum_{\mu=1}^{i-1}(T_{\mu,j+i-\mu} - S_{\mu,j+i-\mu}) + \sum_{\mu=1}^{i}(S_{\mu,j+i-\mu} - T_{\mu,j+i-\mu})$$

$$= T_{ij} + S_{ij} - T_{ij} = S_{ij}$$

if $j - 1 \geq i$, and that

$$X_{ij} + Y_{i,j-1} = S_{ij} + \sum_{\mu=1}^{j-1}(S_{j+i-\mu,\mu} - T_{j+i-\mu,\mu}) + \sum_{\mu=1}^{j-1}(T_{j+i-\mu,\mu} - S_{j+i-\mu,\mu}) = S_{ij}$$

if $j - 1 < i$. This shows $X \boxplus\!\!\!\rightarrow Y = S$ and concludes the proof.   □

Appendix B. Proof of Theorem 5.3.

Proof. Note that (5.8) is equivalent to $X \boxplus\!\!\!\rightarrow Y = S$ and $X \boxplus\!\!\!\downarrow Y = T$, where $S = (S_{ij})$ and $T = (T_{ji})$ are block $k \times k$ matrices such that

(B.1) $S_{ij} = v_i A_{k+1-j}, \quad T_{ji} = w_i A_{k+1-j}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, k+1.$

Then by Proposition 5.2, X and Y satisfying (5.8) exist if and only if

$$(\text{B.2}) \qquad w_j A_{k+1-j} + \sum_{\mu=1}^{j-1} \left(w_{2j-\mu} A_{k+1-\mu} - v_\mu A_{k+1-2j+\mu} \right)$$

$$= v_j A_{k+1-j} + \sum_{\mu=1}^{j-1} \left(v_{2j-\mu} A_{k+1-\mu} - w_\mu A_{k+1-2j+\mu} \right)$$

and

$$(\text{B.3}) \quad \sum_{\mu=1}^{j} \left(v_\mu A_{k-2j+\mu} - w_{2j+1-\mu} A_{k+1-\mu} \right) = \sum_{\mu=1}^{j} \left(w_\mu A_{k-2j+\mu} - v_{2j+1-\mu} A_{k+1-\mu} \right)$$

for $j = 1, \ldots, k$. (Here $v_0 := w_0 := 0$, and for $\mu < 0$ or $\mu > k$, $v_\mu := w_\mu := 0$ and $A_\mu := 0 \in \mathbb{F}^{n \times n}$.) Hence, it is sufficient to prove the statement

$$v = w \quad \Longleftrightarrow \quad \text{(B.2) and (B.3) are satisfied.}$$

"$\Rightarrow$": If $v = w$, then (B.2) and (B.3) are obviously true.

"$\Leftarrow$": We show $v_m = w_m$ for $m = 1, \ldots, k$ by induction on m.

$\boxed{m = 1}$: (B.2) for $j = 1$ yields $v_1 A_k = w_1 A_k$. Since $A_k \neq 0$, this implies $v_1 = w_1$.

$\boxed{m = 2}$: (B.3) for $j = 1$ yields $v_1 A_{k-1} - w_2 A_k = w_1 A_{k-1} - v_2 A_k$. Since $v_1 = w_1$ and $A_k \neq 0$, this implies $v_2 = w_2$.

$\boxed{m - 1 \Rightarrow m}$: Assume first that m is odd, so that $m = 2j - 1$ for some $j \geq 2$. Since by the induction hypothesis we have $v_i = w_i$ for $i = 1, \ldots, 2j - 2$, we obtain from (B.2) that $w_{2j-1} A_k = v_{2j-1} A_k$. This implies $w_{2j-1} = v_{2j-1}$ because $A_k \neq 0$. Next assume that m is even, i.e., $m = 2j$ for some $j \geq 2$. Again, since $v_i = w_i$ for $i = 1, \ldots, 2j - 1$ by the induction hypothesis, we obtain from (B.3) that $w_{2j} A_k = v_{2j} A_k$. This implies $w_{2j} = v_{2j}$ because $A_k \neq 0$.

This concludes the induction. Hence we have $v = w$.

The uniqueness of X and Y and the formulas (5.9) and (5.10) follow directly from Proposition 5.2, the formulas (5.6) and (5.7), and (B.1). $\quad\square$

REFERENCES

[1] E. N. ANTONIOU AND S. VOLOGIANNIDIS, *A new family of companion forms of polynomial matrices*, Electron. J. Linear Algebra, 11 (2004), pp. 78–87.

[2] E. EICH-SOELLNER AND C. FÜHRER, *Numerical Methods in Multibody Systems*, B. G. Teubner, Stuttgart, Germany, 1998.

[3] R. W. FREUND, *Krylov-subspace methods for reduced-order modeling in circuit simulation*, J. Comput. Appl. Math., 123 (2000), pp. 395–421.

[4] I. GOHBERG, M. A. KAASHOEK, AND P. LANCASTER, *General theory of regular matrix polynomials and band Toeplitz operators*, Integral Equations Operator Theory, 11 (1988), pp. 776–882.

[5] I. Gohberg, P. Lancaster, and L. Rodman, *Matrix Polynomials*, Academic Press, New York, 1982.

[6] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.

[7] N. J. Higham, D. S. Mackey, N. Mackey, and F. Tisseur, *Symmetric Linearizations for Matrix Polynomials*, MIMS EPrint 2005.25, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2005.

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

[9] P. Lancaster, *Lambda-Matrices and Vibrating Systems*, Pergamon Press, Oxford, UK, 1966.

[10] P. Lancaster and P. Psarrakos, *A Note on Weak and Strong Linearizations of Regular Matrix Polynomials*, Numerical Analysis Report No. 470, Manchester Centre for Computational Mathematics, Manchester, UK, 2005.

[11] D. S. Mackey, *The characteristic polynomial of a partitioned matrix*, in Linear Algebra Gems, D. Carlson, C. R. Johnson, D. C. Lay, and A. D. Porter, eds., MAA Notes #59, Mathematical Association of America, Washington, DC, 2002, pp. 13–14.

[12] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann, *Palindromic Polynomial Eigenvalue Problems: Good Vibrations from Good Linearizations*, Technical Report 239, DFG Research Center Matheon, "Mathematics for key technologies" in Berlin, TU Berlin, Berlin, Germany, 2005; available online at http://www.matheon.de/.

[13] V. Mehrmann and C. Shi, *Analysis of Higher Order Linear Differential-Algebraic Systems*, Preprint 2004/17, Institut für Mathematik, TU Berlin, D-10623 Berlin, Germany, 2004; available online from http://www.math.tu-berlin.de/preprints/.

[14] V. Mehrmann and D. Watkins, *Structure-preserving methods for computing eigenpairs of large sparse skew-Hamiltonian/Hamiltonian pencils*, SIAM J. Sci. Comput., 22 (2001), pp. 1905–1925.

[15] V. Mehrmann and D. Watkins, *Polynomial eigenvalue problems with Hamiltonian structure*, Electron. Trans. Numer. Anal., 13 (2002), pp. 106–118.

[16] W. Schiehlen, *Advanced Multibody System Dynamics*, Kluwer Academic Publishers, Stuttgart, Germany, 1993.

[17] F. Schmidt, T. Friese, L. Zschiedrich, and P. Deuflhard, *Adaptive Multigrid Methods for the Vectorial Maxwell Eigenvalue Problem for Optical Waveguide Design*, in Mathematics. Key Technology for the Future, W. Jäger and H.-J. Krebs, eds., Springer-Verlag, 2003, pp. 279–292.

[18] C. Shi, *Linear Differential-Algebraic Equations of Higher-Order and the Regularity or Singularity of Matrix Polynomials*, Ph.D. thesis, Institut für Mathematik, Technische Universität, Berlin, 2004.

[19] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, New York, 1990.

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

[21] F. Tisseur and K. Meerbergen, *The quadratic eigenvalue problem*, SIAM Rev., 43 (2001), pp. 235–286.

[22] B. L. van der Waerden, *Modern Algebra*, vol. 1, 2nd ed., Frederick Ungar Publishing, New York, 1953.