Möbius Transformations of Matrix Polynomials

D. Steven Mackey*   Niloufer Mackey*

Christian Mehl†   Volker Mehrmann†

May 20, 2014

Dedicated to Leiba Rodman on the occasion of his 65th birthday

Abstract

We discuss Möbius transformations for general matrix polynomials over arbitrary fields, analyzing their influence on regularity, rank, determinant, constructs such as compound matrices, and on structural features including sparsity and symmetry. Results on the preservation of spectral information contained in elementary divisors, partial multiplicity sequences, invariant pairs, and minimal indices are presented. The effect on canonical forms such as Smith forms and local Smith forms, on relationships of strict equivalence and spectral equivalence, and on the property of being a linearization or quadratification are investigated. We show that many important transformations are special instances of Möbius transformations, and analyze a Möbius connection between alternating and palindromic matrix polynomials. Finally, the use of Möbius transformations in solving polynomial inverse eigenproblems is illustrated.

Key words. Möbius transformation, generalized Cayley transform, matrix polynomial, matrix pencil, Smith form, local Smith form, elementary divisors, partial multiplicity sequence, Jordan characteristic, Jordan chain, invariant pair, compound matrices, minimal indices, minimal bases, structured linearization, palindromic matrix polynomial, alternating matrix polynomial.

AMS subject classification. 65F15, 15A18, 15A21, 15A54, 15A57

1 Introduction

The fundamental role of functions of the form $f(z) = (az + b)/(cz + d)$ in the theory of analytic functions of a complex variable is well-established and classical. Such functions are variously known as fractional linear rational functions [1, 17, 58], bilinear transformations [7, 54, 55, 56] or more commonly, Möbius functions. A particularly important example is the Cayley transformation, see e.g., [54] or the variant in [28, 50], which extends easily to matrix pencils or polynomials. The Cayley transformation is widely used in many areas, such as in the stability analysis of continuous and discrete-time linear systems [33, 36], in the analysis and numerical solution of discrete-time and continuous-time linear-quadratic optimal control problems [50, 59], and in the analysis of geometric integration methods [34].

The main goal of this paper is to present a careful study of the influence of Möbius transformations on properties of general matrix polynomials over arbitrary fields. These include

regularity, rank, determinant, constructs such as the compounds of matrix polynomials, and structural properties such as sparsity and symmetry. We show when spectral information contained in elementary divisors, partial multiplicity sequences, invariant pairs, minimal indices, and minimal bases is preserved, or how its change can be tracked. We study the effect on canonical forms such as Smith forms and local Smith forms, on the relations of strict equivalence and spectral equivalence, and on the property of being a linearization or quadratification. Many of the results presented here are fundamental in that they hold for all matrix polynomials, regular and singular, square and rectangular.

A variety of transformations exploited in the literature [2, 16, 18, 29, 35, 44, 45, 46, 47, 51] will be seen to be special instances of Möbius transformations. The broader theory we present here generalizes and unifies results that were hitherto observed for particular transformations, and provides a more versatile tool for investigating fundamental aspects of matrix polynomials. Important applications include determining the relationships between various classes of structured matrix polynomials (alternating and palindromic, for example), investigating the definiteness of Hermitian matrix polynomials [2], numerical methods for the solution of structured eigenvalue problems and continuous-time Riccati equations via doubling algorithms, (see e.g., [31, 32, 52] and the references therein), the modeling of quantum entanglement via matrix pencils [16], and the triangularization of matrix polynomials [62]. Our results generalize and unify recent and classical results on how Möbius transformations change the finite and infinite elementary divisors of matrix pencils and matrix polynomials [5, 15, 65, 66]; see also [53] for an extension to more general rational transformations. We note that Möbius transformations are also used to study proper rational matrix-valued functions and the Smith-McMillan form [4, 6, 24, 64], but we do not discuss this topic here.

After introducing some definitions and notation in Section 2, Möbius transformations are defined and their fundamental properties established in Section 3. We then investigate the behavior of the Smith form and Jordan characteristic of a matrix polynomial under a Möbius transformation in Sections 4 and 5. The effect of Möbius transformations on invariant pairs is studied in Section 6, on minimal indices and minimal bases in Section 7, and on linearizations and quadratifications of matrix polynomials in Section 8. Section 9 discusses the preservation of sparsity patterns, realization theorems, and the Möbius connection between alternating and palindromic matrix polynomials.

2 Preliminaries

We use $\mathbb{N}$ to denote the set of non-negative integers, $\mathbb{F}$ for an arbitrary field, $\mathbb{F}[\lambda]$ for the ring of polynomials in one variable with coefficients from the field $\mathbb{F}$, and $\mathbb{F}(\lambda)$ for the field of rational functions over $\mathbb{F}$.

A matrix polynomial of grade k has the form $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$, where $A_0, \ldots, A_k \in \mathbb{F}^{m \times n}$. Here we allow any of the coefficient matrices, including A_k, to be the zero matrix. In contrast to the *degree* of a nonzero matrix polynomial, which retains its usual meaning as the largest integer j such that the coefficient of λ^j in $P(\lambda)$ is nonzero, the *grade* indicates that the polynomial $P(\lambda)$ is to be interpreted as an element of the $\mathbb{F}$-vector space of all matrix polynomials of degree less than or equal to k, equivalently, of all $m \times n$ matrix polynomials of grade k. Matrix polynomials that are considered with respect to grade, will be called *graded matrix polynomials*. The notion of grade is crucial in the investigation of spectra of matrix polynomials. As an example consider the matrix with polynomial entries

$$Q(\lambda) = \left[\begin{array}{cc} \lambda + 1 & 0 \\ 0 & \lambda + 1 \end{array} \right].$$

Viewed as a matrix polynomial of grade one or, in other words, as a *matrix pencil*, $Q(\lambda) = \lambda I_2 + I_2$ has a nonsingular leading coefficient and thus does not have infinite eigenvalues. By

contrast, viewing $Q(\lambda) = \lambda^2 0 + \lambda I_2 + I_2$ as a matrix polynomial of grade two, the leading coefficient is now singular and infinity is among the eigenvalues of $Q(\lambda)$. Therefore, throughout this paper a matrix polynomial P must always be accompanied by a *choice* of grade, denoted $\mathrm{grade}(P)$. When the grade is not explicitly specified, then it is to be understood that any choice of grade will suffice.

A polynomial $P(\lambda)$ is said to be *regular* if it is square and invertible when viewed as a matrix over $\mathbb{F}(\lambda)$, equivalently if $\det P(\lambda) \not\equiv 0$; otherwise it is said to be *singular*. The rank of $P(\lambda)$, sometimes called the *normal rank*, is the rank of $P(\lambda)$ when viewed as a matrix with entries in the field $\mathbb{F}(\lambda)$, or equivalently, the size of the largest nonzero minor of $P(\lambda)$.

2.1 Compound matrices and their properties

For references on compound matrices, see [37, Section 0.8], [49, Chapter I.2.7], [57, Section 2 and 28]. We use a variation of the notation in [37] for submatrices of an $m \times n$ matrix A. Let $\eta \subseteq \{1, \ldots, m\}$ and $\kappa \subseteq \{1, \ldots, n\}$ be arbitrary ordered index sets of cardinality $1 \leq j \leq \min(m, n)$. Then $A_{\eta\kappa}$ denotes the $j \times j$ submatrix of A in rows η and columns κ, and the $\eta\kappa$-*minor of order j* of A is $\det A_{\eta\kappa}$. Note that A has $\binom{m}{j} \cdot \binom{n}{j}$ minors of order j.

Definition 2.1 (Compound matrices).
Let A be an $m \times n$ matrix with entries in an arbitrary commutative ring, and let $\ell \leq \min(m, n)$ be a positive integer. Then the ℓth compound matrix (or the ℓth adjugate) of A, denoted by $\mathcal{C}_\ell(A)$, is the $\binom{m}{\ell} \times \binom{n}{\ell}$ matrix whose (η, κ)-entry is the $\ell \times \ell$ minor $\det A_{\eta\kappa}$ of A. Here, the index sets $\eta \subseteq \{1, \ldots, m\}$ and $\kappa \subseteq \{1, \ldots, n\}$ of cardinality ℓ are ordered lexicographically.

Observe that we always have $\mathcal{C}_1(A) = A$, and, if A is square, $\mathcal{C}_n(A) = \det A$. Basic properties of $\mathcal{C}_\ell(A)$ that we need are collected in the next theorem.

Theorem 2.2 (Properties of compound matrices).
Let A be an $m \times n$ matrix with entries in a commutative ring $\mathbb{K}$, and let $\ell \leq \min(m, n)$ be a positive integer. Then

(a) $\mathcal{C}_\ell(A^T) = \big(\mathcal{C}_\ell(A)\big)^T$;

(b) $\mathcal{C}_\ell(\mu A) = \mu^\ell \, \mathcal{C}_\ell(A)$, where $\mu \in \mathbb{K}$;

(c) $\det \mathcal{C}_\ell(A) = (\det A)^\beta$, where $\beta = \binom{n-1}{\ell-1}$, provided that $m = n$;

(d) $\mathcal{C}_\ell(AB) = \mathcal{C}_\ell(A) \, \mathcal{C}_\ell(B)$, provided that $B \in \mathbb{K}^{n \times p}$ and $\ell \leq \min(m, n, p)$;

(e) if A is a diagonal matrix, then so is $\mathcal{C}_\ell(A)$.

When the $m \times n$ matrix polynomial $P(\lambda)$ has grade k, our convention will be that its ℓth compound $\mathcal{C}_\ell(P(\lambda))$ has grade $k\ell$. This is because the degree of the ℓth compound of P can be at most $k\ell$, so the smallest a priori choice for its grade that is guaranteed to work is $k\ell$. In particular, when $m = n$, the scalar polynomial $\det(P(\lambda))$ will have grade kn, since this determinant is identical to $\mathcal{C}_n(P(\lambda))$. In general, it will be advantageous to view $\mathcal{C}_\ell$ as a function from the vector space of $m \times n$ matrix polynomials of grade k to the vector space of $\binom{m}{\ell} \times \binom{n}{\ell}$ matrix polynomials of grade $k\ell$. This will become clear in Section 3.4 when we investigate the effect of Möbius transformations on compounds of matrix polynomials.

3

3 Möbius Transformations

In complex analysis, it is useful to define Möbius functions not just on $\mathbb{C}$, but on the extended complex plane $\mathbb{C} \cup \{\infty\}$, which can be thought of as the Riemann sphere or the complex projective line. We begin therefore with a brief development of $\mathbb{F} \cup \{\infty\} := \mathbb{F}_\infty$, where $\mathbb{F}$ is an arbitrary field. The construction parallels that of $\mathbb{C} \cup \{\infty\}$, and is included here for the convenience of the reader.

On the punctured plane $\mathbb{F}^2 \setminus \{(0,0)\} =: \widehat{\mathbb{F}}^2$ define an equivalence relation: $(e,f) \sim (g,h)$ if $(e,f) = s(g,h)$ for some nonzero scalar $s \in \mathbb{F}$, equivalently if $eh = fg$. Elements of the quotient space $\widehat{\mathbb{F}}^2 / \sim$ can be bijectively associated with the 1-dimensional subspaces of $\mathbb{F}^2$. This quotient space is often referred to as the "projective line" over $\mathbb{F}$, and the mapping

$$
\begin{aligned}
\widehat{\mathbb{F}}^2 &\longrightarrow \mathbb{F}_\infty \\
\begin{bmatrix} e \\ f \end{bmatrix} &\longmapsto \begin{cases} e/f \in \mathbb{F} & \text{if } f \neq 0 \\ \infty & \text{if } f = 0 \,. \end{cases}
\end{aligned}
$$

induces a well-defined bijection $\phi : \widehat{\mathbb{F}}^2 / \sim \; \longrightarrow \; \mathbb{F}_\infty$.

3.1 Möbius Functions over Arbitrary Fields $\mathbb{F}$

It is a classical result that Möbius functions can be characterized via the action of 2×2 matrices [12, 63] with entries in $\mathbb{F}$. When $A \in \mathbb{F}^{2\times 2}$ is nonsingular, i.e., in $GL(2,\mathbb{F})$, its action on $\mathbb{F}^2$ can be naturally viewed as mapping 1-dimensional subspaces of $\mathbb{F}^2$ to one another, and hence as mapping elements of $\mathbb{F}_\infty$ to one another. This can be formally expressed by the composition

$$
\mathbb{F}_\infty \xrightarrow{\phi^{-1}} \widehat{\mathbb{F}}^2 / \sim \xrightarrow{A} \widehat{\mathbb{F}}^2 / \sim \xrightarrow{\phi} \mathbb{F}_\infty \,, \tag{3.1}
$$

which we denote by $\mathsf{m}_A(\lambda)$. One can show that m_A is well-defined for *all* $\lambda \in \mathbb{F}_\infty$ if and only if A is nonsingular. Therefore we will restrict our attention to nonsingular matrices A, and use the phrase "Möbius function" only for functions m_A induced by such A, giving us the following definition.

Definition 3.1. *Let $\mathbb{F}$ be an arbitrary field, and $A \in GL(2,\mathbb{F})$. Then the* Möbius function *on $\mathbb{F}_\infty$ induced by A is the function $\mathsf{m}_A : \mathbb{F}_\infty \to \mathbb{F}_\infty$ defined by the composition (3.1), that is, $\mathsf{m}_A(\lambda) := \phi\big(A\,\phi^{-1}(\lambda)\big)$.*

It immediately follows from the definition that the Möbius functions induced by A and A^{-1} are inverses of each other. We collect some additional useful properties of Möbius functions that follow directly from (3.1).

Proposition 3.2 (Properties of Möbius functions)**.**
Let $A, B \in GL(2,\mathbb{F})$, and I be the 2×2 identity matrix.

(a) m_I *is the identity function on $\mathbb{F}_\infty$.*

(b) $\mathsf{m}_A \circ \mathsf{m}_B = \mathsf{m}_{AB}$.

(c) $(\mathsf{m}_A)^{-1} = \mathsf{m}_{A^{-1}}$.

(d) $\mathsf{m}_{\beta A} = \mathsf{m}_A$, *for any nonzero $\beta \in \mathbb{F}$.*

Properties (a) – (c) of Proposition 3.2 say that the set $\mathcal{M}(\mathbb{F}_\infty)$ of all Möbius functions defined on $\mathbb{F}_\infty$ is a group. Indeed, one can show that the mapping $\psi : GL(2,\mathbb{F}) \to \mathcal{M}(\mathbb{F}_\infty)$ defined by $\psi(A) = \mathsf{m}_A$ is a surjective group homomorphism with $\ker \psi = \{\beta I : \beta \in \mathbb{F}\}$, see [12].

Thus the "Möbius group" $\mathcal{M}(\mathbb{F}_\infty)$ is isomorphic to the quotient group $GL(2, \mathbb{F})/\ker\psi$, which is often referred to as the "projective linear group" $PGL(2, \mathbb{F})$.

Consequently, two matrices induce the same Möbius function if and only if they are scalar multiples of each other. For example, recall that when A is nonsingular, the classical adjoint satisfies $\mathrm{adj}(A) = (\det A)A^{-1}$. For 2×2 matrices, the adjoint is simple to calculate,

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies \mathrm{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

It follows that the Möbius functions associated with A and $\mathrm{adj}(A)$ are inverses of each other.

3.2 Möbius Rational Expressions

It is also useful to be able to work with and manipulate Möbius functions as formal algebraic expressions. In particular, for a matrix A we can view $\mathsf{m}_A(\lambda)$ not just as a function $\mathbb{F}_\infty \to \mathbb{F}_\infty$, but also as the rational expression

$$\mathsf{m}_A(\lambda) \;=\; \frac{a\lambda + b}{c\lambda + d}, \quad \text{where} \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{3.2}$$

Such expressions can be added, multiplied, composed, and simplified in a consistent way by viewing them as formal elements of the field of rational functions $\mathbb{F}(\lambda)$. One can also start with the formal expression (3.2) and use it to generate a function on $\mathbb{F}_\infty$ by invoking standard conventions such as $1/\infty = 0$, $1/0 = \infty$, $(a \cdot \infty + b)/(c \cdot \infty + d) = a/c$, etc. Straightforward manipulations show that the function on $\mathbb{F}_\infty$ thus obtained is the same as the function $\mathsf{m}_A(\lambda) : \mathbb{F}_\infty \to \mathbb{F}_\infty$ in Definition 3.1. This correspondence between rational expressions (3.2) and functions m_A makes it reasonable to use the name *Möbius function* and the notation $\mathsf{m}_A(\lambda)$ interchangeably for both functions and expressions, over arbitrary fields. The context will make clear which interpretation is intended.

As a first example illustrating the use of $\mathsf{m}_A(\lambda)$ as a formal rational expression, we state a simple result that will be needed later. The straightforward proof is omitted.

Lemma 3.3. *Let R be the 2×2 reverse identity matrix*

$$R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{with associated Möbius function} \quad \mathsf{m}_R(\lambda) = \frac{1}{\lambda}.$$

Then for any $A \in GL(2, \mathbb{F})$ we have

$$\mathsf{m}_{RAR}\left(\frac{1}{\lambda}\right) \;=\; \frac{1}{\mathsf{m}_A(\lambda)}. \tag{3.3}$$

3.3 Möbius Transformations of Matrix Polynomials

One of the main motivations for our work is the study of the relationships between different classes of structured matrix polynomials. Clearly such a study can be greatly aided by fashioning transformations that allow results about one structured class to be translated into results about another structured class. Indeed, this has been done using particular Möbius transformations in a number of instances [44, 51]. The development of general Möbius transformations in this section will bring several previously unrelated techniques under one umbrella, and set the stage for the main new results presented in later sections.

Definition 3.4 (Möbius Transformation)**.**
Let V be the vector space of all $m \times n$ matrix polynomials of grade k over the field $\mathbb{F}$, and let

$A \in GL(2, \mathbb{F})$. *Then the* Möbius transformation on V induced by A is the map $\mathbf{M}_A : V \to V$ *defined by*

$$\mathbf{M}_A \left(\sum_{i=0}^k B_i \lambda^i \right) (\mu) = \sum_{i=0}^k B_i \, (a\mu + b)^i (c\mu + d)^{k-i}, \quad \text{where} \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

It is worth pointing out that a Möbius transformation acts on graded polynomials, returning polynomials of the same grade (although the degree may increase, decrease, or stay the same, depending on the polynomial). In fact, $\mathbf{M}_A$ is a linear operator on V.

Proposition 3.5. *For any $A \in GL(2, \mathbb{F})$, $\mathbf{M}_A$ is an $\mathbb{F}$-linear operator on the vector space V of all $m \times n$ matrix polynomials of grade k, that is,*

$$\mathbf{M}_A(P + Q) = \mathbf{M}_A(P) + \mathbf{M}_A(Q), \quad \text{and} \quad \mathbf{M}_A(\beta P) = \beta \, \mathbf{M}_A(P),$$

for any $\beta \in \mathbb{F}$ and for all $P, Q \in V$.

Proof. The proof follows immediately from Definition 3.4. □

As in the case of a Möbius function m_A on $\mathbb{F}_\infty$, a Möbius transformation $\mathbf{M}_A$ on V can also be formally calculated via a rational expression:

$$(\mathbf{M}_A(P))\,(\mu) = (c\mu + d)^k P \left(\frac{a\mu + b}{c\mu + d} \right), \quad \text{where} \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{3.4}$$

or equivalently,

$$(\mathbf{M}_A(P))\,(\mu) = (c\mu + d)^k P\,(\mathsf{m}_A(\mu)). \tag{3.5}$$

We now present several examples. The first example shows that a Möbius transformation can change the degree of the matrix polynomial it acts on. The other examples illustrate how operators previously used in the literature are special instances of Möbius transformations.

Example 3.6. Consider the Möbius transformation

$$\mathbf{M}_A : V \to V, \quad \text{where} \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

and V is the real vector space of all 2×2 matrix polynomials of grade 2. Computing the action of $\mathbf{M}_A$ on three polynomials in V,

$$P = \lambda I, \quad Q = (\lambda^2 + \lambda - 2)I, \quad \text{and} \quad S = \lambda^2 I,$$

we find

$$\mathbf{M}_A(P) = \lambda(\lambda + 1)I, \quad \mathbf{M}_A(Q) = (3\lambda + 1)I, \quad \mathbf{M}_A(S) = (\lambda + 1)^2 I.$$

Thus the degree of an input polynomial can increase, decrease, or stay the same under a Möbius transformation. Note, however, that if the degree and grade of P are equal, then $\deg \mathbf{M}_A(P) \leq \deg P$.

Example 3.7. The *reversal of a (matrix) polynomial* is an important notion used to define infinite elementary divisors and strong linearizations of matrix polynomials [29, 41, 45]. The *reversal* operation rev_k, that reverses the order of the coefficients of a matrix polynomial with respect to grade k, was defined and used in [47]. When viewed as an operator on the vector space of all $m \times n$ matrix polynomials of grade k, rev_k is just the Möbius transformation induced by the 2×2 reverse identity matrix R:

$$R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \implies (\mathbf{M}_R(P))(\mu) = \sum_{i=0}^k B_i \mu^{k-i} = \mu^k P \left(\frac{1}{\mu} \right) = (\mathrm{rev}_k P)(\mu).$$

This can be seen directly from Definition 3.4, or from the rational expression (3.4).

Example 3.8. The map $P(\lambda) \mapsto P(-\lambda)$ used in defining T-even and T-odd matrix polynomials [46] is a Möbius transformation, induced by the matrix $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$.

Example 3.9. Any translation $P(\lambda) \mapsto P(\lambda + b)$ where $b \neq 0$, is a Möbius transformation induced by the matrix $\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}$.

Example 3.10. The classical *Cayley transformation* is a useful device to convert one matrix structure (e.g., skew-Hermitian or Hamiltonian) into another (unitary or symplectic, respectively). It was extended from matrices to pencils [42, 51], and then generalized to matrix polynomials in [44]. Observe that the Möbius transformations induced by

$$A_{+1} \; = \; \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad A_{-1} \; = \; \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \tag{3.6}$$

are the Cayley transformations

$$\mathcal{C}_{+1}(P)(\mu) := (1 - \mu)^k P\left(\frac{1+\mu}{1-\mu}\right) \quad \text{and} \quad \mathcal{C}_{-1}(P)(\mu) := (\mu + 1)^k P\left(\frac{\mu - 1}{\mu + 1}\right), \tag{3.7}$$

respectively, introduced in [44], where they were used to relate palindromic with T- even and T-odd matrix polynomials.

Example 3.11. Recall that one of the classical Cayley transforms of a square matrix S is $\widetilde{S} := (I + S)(I - S)^{-1}$, assuming $I - S$ is nonsingular. This can be viewed as formally substituting S for λ in the Möbius function $\mathsf{m}_{A_{+1}}(\lambda) = \frac{\lambda+1}{-\lambda+1}$, or as evaluating the function $\mathsf{m}_{A_{+1}}$ at the matrix S. Now if we consider the Möbius transformation $\mathbf{M}_{A_{-1}}$ applied to the *pencil* naturally associated with the matrix S, i.e., $\lambda I - S$, then we obtain

$$\mathbf{M}_{A_{-1}}(\lambda I - S) \; = \; (\lambda + 1)\big[\mathsf{m}_{A_{-1}}(\lambda)I - S\big] \; = \; (\lambda - 1)I - (\lambda + 1)S \; = \; \lambda(I - S) - (I + S).$$

When $I - S$ is nonsingular, this pencil is strictly equivalent (i.e., under pre- and post-multiplication with nonsingular constant matrices) to the pencil $\lambda I - \widetilde{S}$ naturally associated with the classical Cayley transform $\widetilde{S} = \mathsf{m}_{A_{+1}}(S)$. Thus the *matrix* Cayley transform associated with A_{+1} is intimately connected to the *pencil* Cayley transformation associated with A_{-1}.

More generally, for any matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL(2, \mathbb{F})$, we define $\mathsf{m}_A(S)$ to be the matrix naturally obtained from the formal expression $\frac{aS+b}{cS+d}$, i.e.,

$$\mathsf{m}_A(S) := (aS + bI)(cS + dI)^{-1}, \tag{3.8}$$

as long as $(cS + dI)$ is invertible. A straightforward calculation shows that the pencils

$$\mathbf{M}_A(\lambda I - S) \quad \text{and} \quad \lambda I - \mathsf{m}_{A^{-1}}(S) \tag{3.9}$$

are strictly equivalent, as long as the matrix $(aI - cS)$ is nonsingular.

Example 3.12. Möbius transformations $\mathbf{M}_A$ induced by rotation matrices

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{for some} \quad \theta \in \mathbb{R}$$

were called *homogeneous rotations* in [2, 35], where they played an important role in the analysis and classification of Hermitian matrix polynomials with spectrum contained in the extended real line $\mathbb{R}_\infty$.

Example 3.13. To facilitate the investigation of the pseudospectra of matrix polynomials, [18] introduces and exploits a certain Möbius transformation named *reversal with respect to a Lagrange basis.*

Example 3.14. The phenomenon of *quantum entanglement* is modelled in [16] using matrix pencils. In this model the properties of Möbius transformations of pencils play a key role in understanding the equivalence classes of quantum states in tripartite quantum systems.

Several basic properties of Möbius transformations follow immediately from Definition 3.4 and the following simple consequence of that definition — any Möbius transformation acts *entry-wise* on a matrix polynomial $P(\lambda)$, in the sense that

$$\left[\mathbf{M}_A(P)\right]_{ij} \;=\; \mathbf{M}_A\big(P_{ij}\big). \tag{3.10}$$

Here it is to be understood that the scalar polynomial $P_{ij}(\lambda)$ has the same grade k as its parent polynomial $P(\lambda)$, and the Möbius transformations $\mathbf{M}_A$ in (3.10) are both taken with respect to that same grade k. This observation extends immediately to arbitrary submatrices, so that

$$\left[\mathbf{M}_A(P)\right]_{\eta\kappa} \;=\; \mathbf{M}_A\big(P_{\eta\kappa}\big), \tag{3.11}$$

for any row and column index sets η and κ. As a consequence it is easy to see that any $\mathbf{M}_A$ is compatible with direct sums, transposes, and conjugation, using the following definitions and conventions.

Definition 3.15. *Suppose P and Q are matrix polynomials (not necessarily of the same size) over an arbitrary field $\mathbb{F}$, and $P(\lambda) = \sum_{j=0}^{k} B_j \lambda^j$. Then*

(a) *the transpose of P is the polynomial $P^T(\lambda) := \sum_{j=0}^{k} B_j^T \lambda^j$,*

(b) *for the field $\mathbb{F} = \mathbb{C}$, the conjugate of P is the polynomial $\overline{P}(\lambda) := \sum_{j=0}^{k} \overline{B}_j \lambda^j$, and hence the conjugate transpose of P is $P^*(\lambda) = \sum_{j=0}^{k} B_j^* \lambda^j$,*

(c) *if P and Q both have grade k, then $P \oplus Q := \mathrm{diag}(P, Q)$ has the same grade k.*

Proposition 3.16 (Properties of Möbius transformations).
Let P and Q be any two matrix polynomials (not necessarily of the same size) of grade k over an arbitrary field $\mathbb{F}$, and let $A \in GL(2, \mathbb{F})$. Then

(a) $\mathbf{M}_A(P^T) \;=\; \big(\mathbf{M}_A(P)\big)^T$,

(b) *if $\mathbb{F} = \mathbb{C}$ and $A \in GL(2, \mathbb{C})$, then $\overline{\mathbf{M}_A(P)} = \mathbf{M}_{\overline{A}}(\overline{P})$ and $\big(\mathbf{M}_A(P)\big)^* = \mathbf{M}_{\overline{A}}(P^*)$,*

(c) $\mathbf{M}_A(P \oplus Q) \;=\; \mathbf{M}_A(P) \oplus \mathbf{M}_A(Q)$.

Proof. The proofs are straightforward, and hence omitted. $\qquad\qquad\square$

Remark 3.17. Note that from Proposition 3.16(b) it follows that Hermitian structure of a matrix polynomial is preserved by any Möbius transformation induced by a real matrix. In other words, when $P^* = P$ and $A \in GL(2, \mathbb{R})$, then $(\mathbf{M}_A(P))^* = \mathbf{M}_A(P)$. Over an arbitrary field $\mathbb{F}$, it follows immediately from Proposition 3.16(a) and Proposition 3.5 that symmetric and skew-symmetric structure is preserved by any Möbius transformation. That is, when $P^T = \pm P$ and $A \in GL(2, \mathbb{F})$, then $(\mathbf{M}_A(P))^T = \pm \mathbf{M}_A(P)$.

We saw in Section 3.1 that the collection of all Möbius functions on $\mathbb{F}_\infty$ forms a group that is isomorphic to a quotient group of $GL(2, \mathbb{F})$. It is then natural to ask whether an analogous result holds for the set of Möbius transformations on matrix polynomials.

Theorem 3.18 (Further properties of Möbius transformations).
Let V be the vector space of all $m \times n$ polynomials of grade k, over an arbitrary field $\mathbb{F}$. Let $A, B \in GL(2, \mathbb{F})$, and let I be the 2×2 identity matrix. Then

(a) $\mathbf{M}_I$ *is the identity operator on* V.

(b) $\mathbf{M}_A \circ \mathbf{M}_B = \mathbf{M}_{BA}$.

(c) $(\mathbf{M}_A)^{-1} = \mathbf{M}_{A^{-1}}$, *so* $\mathbf{M}_A$ *is a bijection on* V.

(d) $\mathbf{M}_{\beta A} = \beta^k \mathbf{M}_A$, *for any* $\beta \in \mathbb{F}$.

Proof. Parts (a) and (d) are immediate from Definition 3.4. Part (c) follows directly from (b) and (a), so all that remains is to demonstrate part (b). Letting

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} e & f \\ g & h \end{bmatrix},$$

we see that

$$
\begin{aligned}
\mathbf{M}_A\big(\mathbf{M}_B(P)\big)(\mu) &= \mathbf{M}_A\Big[(g\lambda + h)^k P\big(\mathsf{m}_B(\lambda)\big)\Big](\mu) \\
&= (c\mu + d)^k \Big[\big(g\mathsf{m}_A(\mu) + h\big)^k P\big(\mathsf{m}_B(\mathsf{m}_A(\mu))\big)\Big] \\
&= \Big[(c\mu + d)\big(g\mathsf{m}_A(\mu) + h\big)\Big]^k P\big(\mathsf{m}_{BA}(\mu)\big) \\
&= \Big[(ga + hc)\mu + (gb + hd)\Big]^k P\big(\mathsf{m}_{BA}(\mu)\big) \\
&= \mathbf{M}_{BA}(P)(\mu),
\end{aligned}
$$

as desired. $\qquad\square$

The properties in Theorem 3.18 say that the set of all Möbius transformations on V forms a group under composition. Property (b) shows this group is an anti-homomorphic image of $GL(2, \mathbb{F})$, while (d) implies that A and βA usually do not define the same Möbius transformation, by contrast with Möbius functions (see Proposition 3.2(d)).

We now revisit some of the examples presented earlier in this section, in the light of Theorem 3.18.

Example 3.19. Observe that the matrix R associated with the reversal operator (see Example 3.7) satisfies $R^2 = I$. Thus the property

$$\mathrm{rev}_k\left(\mathrm{rev}_k(P)\right) = P \quad \text{for} \quad k \geq \deg(P)$$

used in [47] can be seen to be a special case of Theorem 3.18(b).

Example 3.20. Notice that the product of the matrices A_{+1} and A_{-1} introduced in (3.6) is just $2I$. Hence by Theorem 3.18,

$$\mathbf{M}_{A_{+1}} \circ \mathbf{M}_{A_{-1}} = \mathbf{M}_{2I} = 2^k(\mathbf{M}_I).$$

When expressed in terms of the associated Cayley transformations, this yields the identity $\mathcal{C}_{+1}(\mathcal{C}_{-1}(P)) = \mathcal{C}_{-1}(\mathcal{C}_{+1}(P)) = 2^k P$, which was derived in [44]. The calculation presented here is both simpler and more insightful.

Example 3.21. Since the set of rotation matrices in Example 3.12 is closed under multiplication, then by Theorem 3.18 so are the corresponding Möbius transformations. Thus the set of all homogeneous rotations forms a group under composition.

Remark 3.22. As illustrated in Example 3.6, a Möbius transformation can alter the degree of its input polynomial. If Möbius transformations were always defined with respect to degree, then the fundamental property in Theorem 3.18(b) would sometimes fail, as an examination of its proof will show. By defining Möbius transformations with respect to grade, we obtain a property of uniform applicability.

We now bring two types of matrix polynomial products into play.

Proposition 3.23 (Möbius transforms of products).
Let P and Q be matrix polynomials of grades k and ℓ, respectively, over an arbitrary field.

(a) *If PQ is defined and of grade $k + \ell$, then $\mathbf{M}_A(PQ) = \mathbf{M}_A(P)\mathbf{M}_A(Q)$.*

(b) *If $P \otimes Q$ has grade $k + \ell$, then $\mathbf{M}_A(P \otimes Q) = \mathbf{M}_A(P) \otimes \mathbf{M}_A(Q)$.*

Proof. (a) Recall that by definition, each Möbius transformation is taken with respect to the grade of the corresponding polynomial. Thus, using (3.5) we have

$$\mathbf{M}_A(P)(\mu) = (c\mu + d)^k P(\mathsf{m}_A(\mu)) \quad \text{and} \quad \mathbf{M}_A(Q)(\mu) = (c\mu + d)^\ell Q(\mathsf{m}_A(\mu)).$$

Taking their product yields

$$
\begin{aligned}
\mathbf{M}_A(P)(\mu) \cdot \mathbf{M}_A(Q)(\mu) &= (c\mu + d)^{k+\ell} P(\mathsf{m}_A(\mu))Q(\mathsf{m}_A(\mu)) \\
&= (c\mu + d)^{k+\ell} PQ(\mathsf{m}_A(\mu)) \\
&= \mathbf{M}_A(PQ)(\mu),
\end{aligned}
\tag{3.12}
$$

as desired.

(b) The proof is formally the same as for part (a); just replace ordinary matrix multiplication everywhere by Kronecker product. $\square$

These results motivate the adoption of two conventions: that the grade of the Kronecker product of any two graded matrix polynomials is equal to the sum of their individual grades, and if those polynomials are conformable for multiplication, then the grade of their ordinary product is also equal to the sum of their individual grades. A subtle point worth mentioning is that the equality in (3.12) can fail if Möbius transformations are taken with respect to degree, rather than grade. When P, Q are *matrix* polynomials (as opposed to scalar ones) of degree k and ℓ respectively, the degree of their product PQ can drop below the sum of the degrees of P and Q. If Möbius transformations were taken only with respect to degree, then we would have instances when

$$
\begin{aligned}
\mathbf{M}_A(PQ)(\mu) &= (c\mu + d)^j PQ(\mathsf{m}_A(\mu)) \quad \text{with } j < k + \ell \\
&\neq \mathbf{M}_A(P)\mathbf{M}_A(Q).
\end{aligned}
$$

In contrast to this, for scalar polynomials we always have equality.

Corollary 3.24 (Multiplicative and divisibility properties).
Let p, q be nonzero scalar polynomials over $\mathbb{F}$, with grades equal to their degrees, and let $A \in GL(2, \mathbb{F})$. Then

(a) $\mathbf{M}_A(pq) = \mathbf{M}_A(p)\mathbf{M}_A(q).$

(b) $p \,|\, q \implies \mathbf{M}_A(p) \,|\, \mathbf{M}_A(q).$

(c) *p is $\mathbb{F}$-irreducible $\implies \mathbf{M}_A(p)$ is $\mathbb{F}$-irreducible.*

(d) *p, q coprime $\implies \mathbf{M}_A(p)$ and $\mathbf{M}_A(q)$ are coprime.*

For each of the implications in (b), (c), and (d), the converse does not hold.

Proof. Part (a) is a special case of Proposition 3.23. For part (b), since $p|q$, there exists a polynomial r such that $q = pr$. Then (a) implies $\mathbf{M}_A(q) = \mathbf{M}_A(p)\mathbf{M}_A(r)$, from which the desired result is immediate.

To prove (c) we show the contrapositive, i.e., if $\mathbf{M}_A(p)$ has a nontrivial factor over $\mathbb{F}$, then so does p. So let

$$\mathbf{M}_A(p) = r(\lambda)s(\lambda) \quad \text{with} \quad 1 \leq \deg r, \deg s < \deg \mathbf{M}_A(p), \tag{3.13}$$

and observe that $\deg \mathbf{M}_A(p) \leq \deg p$ by the remark in Example 3.6. Applying $\mathbf{M}_{A^{-1}}$ with respect to grade $\mathbf{M}_A(p) = \deg p$ (rather than with respect to $\deg \mathbf{M}_A(p)$) to both sides of the equation in (3.13) then yields $p(\lambda) = \mathbf{M}_{A^{-1}}(r(\lambda)s(\lambda))$, or equivalently

$$p(\lambda) = (e\lambda + f)^{\deg p - \deg \mathbf{M}_A(p)}\mathbf{M}_{A^{-1}}(r)\mathbf{M}_{A^{-1}}(s), \tag{3.14}$$

where $(e\lambda + f)$ is the denominator of the Möbius function $\mathsf{m}_{A^{-1}}$, and the transformations $\mathbf{M}_{A^{-1}}$ on the right-hand side of (3.14) are once again taken with respect to degree. Since the first factor on the right-hand side of (3.14) has degree at most $\deg p - \deg \mathbf{M}_A(p)$, we have

$$\deg\big(\mathbf{M}_{A^{-1}}(r)\mathbf{M}_{A^{-1}}(s)\big) \geq \deg \mathbf{M}_A(p) = \deg r + \deg s \geq 2.$$

But $\deg \mathbf{M}_{A^{-1}}(r) \leq \deg r$ and $\deg \mathbf{M}_{A^{-1}}(s) \leq \deg s$ by the remark in Example 3.6, so we must have the *equalities* $\deg \mathbf{M}_{A^{-1}}(r) = \deg r$ and $\deg \mathbf{M}_{A^{-1}}(s) = \deg s$. Thus both $\mathbf{M}_{A^{-1}}(r)$ and $\mathbf{M}_{A^{-1}}(s)$ are nontrivial factors of p, and the proof of (c) is complete.

For part (d) we again show the contrapositive; if $\mathbf{M}_A(p)$ and $\mathbf{M}_A(q)$ have a nontrivial common factor $r(\lambda)$, then from the proof of (c) we know that $\mathbf{M}_{A^{-1}}(r)$ will be a nontrivial common factor for p and q, and we are done.

Finally, it is easy to build counterexamples to show that the converses of (b), (c), and (d) fail to hold. For all three cases take $p(\lambda) = \lambda^2 + \lambda$ and $\mathbf{M}_A = \text{rev}$; for (b) let $q(\lambda) = \lambda^2 - 1$, but for (d) use instead $q(\lambda) = 2\lambda^2 + \lambda$. $\qquad\square$

Remark 3.25. It is worth noting that alternative proofs for parts (c) and (d) of Corollary 3.24 can be fashioned by passing to the algebraic closure $\overline{\mathbb{F}}$ and using Lemma 5.6.

Example 3.26. The multiplicative property of reversals acting on scalar polynomials [47, Lemma 3.9]
$$\text{rev}_{j+\ell}(pq) = \text{rev}_j(p) \cdot \text{rev}_\ell(q)$$
where $j \geq \deg p$, and $\ell \geq \deg q$, is just a special case of Proposition 3.23.

3.4 Interaction of Möbius Transformations with Rank and Compounds

The first step in understanding how Möbius transformations affect rank and compounds of matrix polynomials is to see how they interact with determinants. The next proposition shows that determinants commute with any Möbius transformation.

Proposition 3.27. *Let P be any $n \times n$ matrix polynomial of grade k. Then*

$$\det\big(\mathbf{M}_A(P)\big) = \mathbf{M}_A(\det P) \tag{3.15}$$

for any $A \in GL(2, \mathbb{F})$.

Proof. Recall that $\det(P)$ is a polynomial of grade kn, by the convention established at the end of Section 2.1. Thus the operators $\mathbf{M}_A$ on the two sides of (3.15) operate with respect

to different grades in general, grade k on the left hand side, and grade kn on the right hand side. Then we have

$$\det\big(\mathbf{M}_A(P)\big) \;=\; \det\Big[(c\mu+d)^k P(\mathsf{m}_A(\mu))\Big] \;=\; (c\mu+d)^{kn}\det\big[P(\mathsf{m}_A(\mu))\big]$$
$$= (c\mu+d)^{kn}(\det P)(\mathsf{m}_A(\mu)) \;=\; \mathbf{M}_A(\det P)\,,$$

and the proof is complete. $\qquad\square$

As an immediate corollary we see that regularity of a matrix polynomial is preserved by Möbius transformations.

Corollary 3.28. *Let P be any $n \times n$ matrix polynomial of grade k, and let $A \in GL(2,\mathbb{F})$. Then P is regular if and only if $\mathbf{M}_A(P)$ is regular.*

More generally, Möbius transformations preserve the normal rank of a matrix polynomial.

Proposition 3.29. *Let P be any $m \times n$ matrix polynomial of grade k. Then for any $A \in GL(2,\mathbb{F})$,*
$$\operatorname{rank}\mathbf{M}_A(P) \;=\; \operatorname{rank}(P).$$

Proof. Consider a general $\ell \times \ell$ submatrix $P_{\eta\kappa}$ of P and the corresponding submatrix $\big(\mathbf{M}_A(P)\big)_{\eta\kappa}$ of $\mathbf{M}_A(P)$, where η and κ are row and column index sets, respectively, of cardinality ℓ.

Since $\mathbf{M}_A$ acts entry-wise, by (3.11) we have $\big(\mathbf{M}_A(P)\big)_{\eta\kappa} = \mathbf{M}_A(P_{\eta\kappa})$. Taking determinants and then using Proposition 3.27 we get

$$\det\big(\mathbf{M}_A(P)\big)_{\eta\kappa} = \det\big(\mathbf{M}_A(P_{\eta\kappa})\big) = \mathbf{M}_A\big(\det(P_{\eta\kappa})\big). \tag{3.16}$$

Thus by Theorem 3.18(c) we have

$$\det\big(\mathbf{M}_A(P)\big)_{\eta\kappa} = 0 \;\;\Leftrightarrow\;\; \mathbf{M}_A\big(\det(P_{\eta\kappa})\big) = 0 \;\;\Leftrightarrow\;\; \det(P_{\eta\kappa}) = 0\,,$$

and we conclude that nonsingular submatrices (i.e., matrices that are nonsingular over the field $\mathbb{F}(\lambda)$) occur in exactly the same locations in P and $\mathbf{M}_A(P)$. Since rank is the size of the largest nonsingular submatrix, the desired conclusion now follows. $\qquad\square$

We saw earlier in Proposition 3.27 that Möbius transformations and determinants commute. The observation that the determinant of an $n \times n$ matrix is its nth compound prompts one to investigate whether this commuting property holds for other compounds.

Theorem 3.30. *Let P be an $m \times n$ matrix polynomial of grade k, and let $A \in GL(2,\mathbb{F})$. Then*
$$\mathcal{C}_\ell\big(\mathbf{M}_A(P)\big) \;=\; \mathbf{M}_A\big(\mathcal{C}_\ell(P)\big) \quad for \quad \ell = 1, 2, \ldots, \min\{m,n\}. \tag{3.17}$$

Proof. Recall from Section 2.1 that the ℓth compound $\mathcal{C}_\ell(P)$ has grade $k\ell$. This means that in (3.17), the Möbius transformation on the right hand side is with respect to grade $k\ell$, while the one on the left hand side is with respect to grade k. We will establish (3.17) by showing that corresponding entries of the matrices in question are equal.

Consider arbitrary row and column index sets η and κ, respectively, of cardinality ℓ. Then by definition the (η,κ)-entry of $\mathcal{C}_\ell\big(\mathbf{M}_A(P)\big)$ is just $\det\big(\mathbf{M}_A(P)\big)_{\eta\kappa}$, and by (3.16) we have $\det\big(\mathbf{M}_A(P)\big)_{\eta\kappa} = \mathbf{M}_A\big(\det(P_{\eta\kappa})\big)$. Since $\det(P_{\eta\kappa})$ is the (η,κ)-entry of $\mathcal{C}_\ell(P)$, we see that

$$\big(\mathcal{C}_\ell(\mathbf{M}_A(P))\big)_{\eta\kappa} \;=\; \mathbf{M}_A\big((\mathcal{C}_\ell(P))_{\eta\kappa}\big) \;=\; \big(\mathbf{M}_A(\mathcal{C}_\ell(P))\big)_{\eta\kappa}\,,$$

by the observation in (3.10). Thus the matrices on each side in (3.17) agree entry-by-entry, and the proof is complete. $\qquad\square$

We remarked at the end of Section 2.1 that *a priori*, the only sensible choice for the grade of the ℓth compound of a grade k matrix polynomial was $k\ell$. The results of this subsection reinforce that choice, demonstrating that rank, determinant and compounds cohere nicely with Möbius transformations once we use the framework of graded polynomials; and they make a case, in hindsight, for defining the grade of $\det P$ and $\mathcal{C}_\ell(P)$ as kn and $k\ell$ respectively, where k is the grade of P.

4   Jordan Characteristic of Matrix Polynomials

In this section we introduce the Jordan characteristic of a matrix polynomial, develop some tools for calculating this invariant, and see how the Jordan characteristic of a polynomial P is related to that of its reversal $\operatorname{rev} P$.

4.1   Smith Form and Jordan Characteristic

Recall that an $n \times n$ matrix polynomial $E(\lambda)$ is said to be *unimodular* if $\det E(\lambda)$ is a nonzero constant. Two $m \times n$ matrix polynomials $P(\lambda), Q(\lambda)$ are said to be *unimodularly equivalent*, denoted by $P \sim Q$, if there exist unimodular matrix polynomials $E(\lambda)$ and $F(\lambda)$ of size $m \times m$ and $n \times n$, respectively, such that

$$Q(\lambda) \;=\; E(\lambda)P(\lambda)F(\lambda)\,. \tag{4.1}$$

If $E(\lambda)$ and $F(\lambda)$ in (4.1) are nonsingular constant matrices, then P and Q are said to be *strictly equivalent*.

Theorem 4.1 (Smith form (Frobenius, 1878)[26])**.**
Let $P(\lambda)$ be an $m \times n$ matrix polynomial over an arbitrary field $\mathbb{F}$. Then there exists $r \in \mathbb{N}$, and unimodular matrix polynomials $E(\lambda)$ and $F(\lambda)$ over $\mathbb{F}$ of size $m \times m$ and $n \times n$, respectively, such that

$$E(\lambda)P(\lambda)F(\lambda) = \operatorname{diag}(d_1(\lambda), \ldots, d_{\min\{m,n\}}(\lambda)) =: D(\lambda),$$

where $d_{r+1}(\lambda), \ldots, d_{\min\{m,n\}}(\lambda)$ are identically zero, while $d_1(\lambda), \ldots, d_r(\lambda)$ are monic and satisfy the divisibility chain property, i.e., $d_j(\lambda)$ is a divisor of $d_{j+1}(\lambda)$ for $j = 1, \ldots, r-1$. Moreover, $D(\lambda)$ is unique.

The nonzero diagonal elements $d_j(\lambda)$, $j = 1, \ldots, r$ in the Smith form of $P(\lambda)$ are called the *invariant factors* or *invariant polynomials* of $P(\lambda)$.

Observe that the uniqueness of the Smith form over a field $\mathbb{F}$ implies that the Smith form is insensitive to field extensions. In particular, the Smith form of P over $\mathbb{F}$ is the same as that over $\overline{\mathbb{F}}$, the algebraic closure of $\mathbb{F}$. It will sometimes be more convenient to work over $\overline{\mathbb{F}}$, where each invariant polynomial can be completely decomposed into a product of linear factors.

Definition 4.2 (Partial multiplicity sequences)**.**
Let $P(\lambda)$ be an $m \times n$ matrix polynomial over a field $\mathbb{F}$, with $\operatorname{rank}(P) = r$ and $\operatorname{grade}(P) = k$. For any $\lambda_0 \in \mathbb{F}$, the invariant polynomials $d_i(\lambda)$ of P for $1 \le i \le r$ can each be uniquely factored as

$$d_i(\lambda) \;=\; (\lambda - \lambda_0)^{\alpha_i}\, p_i(\lambda) \quad \text{with} \quad \alpha_i \in \mathbb{N}, \ p_i(\lambda_0) \neq 0\,.$$

The sequence of exponents $(\alpha_1, \alpha_2, \ldots, \alpha_r)$, which satisfies the condition $0 \le \alpha_1 \le \alpha_2 \le \cdots \le \alpha_r$ by the divisibility chain property of the Smith form, is called the partial multiplicity sequence of P at λ_0, denoted $\mathcal{J}(P, \lambda_0)$. For $\lambda_0 = \infty$, the partial multiplicity sequence $\mathcal{J}(P, \infty)$ is defined to be identical with $\mathcal{J}(\operatorname{rev}_k P, 0) = \mathcal{J}(\mathbf{M}_R(P), 0)$, where R is as in Example 3.7.

Note that since $\operatorname{rank}(\operatorname{rev}_k P) = \operatorname{rank}(P) = r$ by Proposition 3.29, we see that the sequence $\mathcal{J}(P, \infty)$ is also of length r.

Remark 4.3. Note that the sequence $\mathcal{J}(P, \lambda_0)$ may consist of all *zeroes*; indeed, this occurs for all but a finite subset of values $\lambda_0 \in \mathbb{F}_\infty$. An *eigenvalue* of P is an element $\lambda_0 \in \mathbb{F}_\infty$ such that $\mathcal{J}(P, \lambda_0)$ does *not* consist of all zeroes. The *spectrum* of P, denoted by $\sigma(P)$, is the collection of all the eigenvalues of P. If an eigenvalue λ_0 has $\mathcal{J}(P, \lambda_0) = (\alpha_1, \alpha_2, \ldots, \alpha_r)$, then the *algebraic multiplicity* of λ_0 is just the sum $\alpha_1 + \alpha_2 + \cdots + \alpha_r$, while the *geometric multiplicity* of λ_0 is the number of positive terms α_j in $\mathcal{J}(P, \lambda_0)$. The *elementary divisors* associated with a finite λ_0 are the collection of all powers $\{(\lambda - \lambda_0)^{\alpha_j} : \alpha_j > 0\}$, including repetitions, while the *elementary divisors* associated with the eigenvalue ∞ are the elementary divisors of $\operatorname{rev}_k P$ associated with the eigenvalue 0, where the reversal is taken with respect to the grade k of P.

It is worth stressing the importance of viewing the partial multiplicities of a fixed λ_0 as a *sequence*. In a number of situations, especially for matrix polynomials with structure [46, 47, 48], it is essential to consider certain *subsequences* of partial multiplicities, which can be subtly constrained by the matrix polynomial structure. Indeed, even the zeroes in the partial multiplicity sequences of structured matrix polynomials can sometimes have nontrivial significance [46, 47, 48].

In the next definition we gather together *all* of the partial multiplicity sequences of P (one for each element of $\mathbb{F}_\infty$) into a single object called the Jordan characteristic of P.

Definition 4.4 (Jordan characteristic of a matrix polynomial).
Suppose $P(\lambda)$ is an $m \times n$ matrix polynomial over a field $\mathbb{F}$ with $\operatorname{rank}(P) = r$. Let $\mathbb{N}_{\leq}^r$ denote the space of all ordered sequences of natural numbers of length r, i.e.,

$$\mathbb{N}_{\leq}^r := \left\{ (\alpha_1, \alpha_2, \ldots, \alpha_r) : \alpha_j \in \mathbb{N} \ \text{and} \ 0 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_r \right\}.$$

Then the Jordan characteristic of P over the field $\mathbb{F}$ is the mapping $\mathcal{J}(P, \mathbb{F})$ defined by

$$\begin{aligned} \mathcal{J}(P, \mathbb{F}) : \quad \mathbb{F}_\infty &\longrightarrow \quad \mathbb{N}_{\leq}^r \\ \lambda_0 &\longmapsto \quad \mathcal{J}(P, \lambda_0) \end{aligned} \tag{4.2}$$

If the field $\mathbb{F}$ is not algebraically closed, then we denote by $\mathcal{J}(P, \overline{\mathbb{F}})$ the Jordan characteristic of P over the algebraic closure $\overline{\mathbb{F}}$. When the field $\mathbb{F}$ is clear from context, then for simplicity we use the more concise notation $\mathcal{J}(P)$.

There are a number of useful properties of Jordan characteristic that follow almost immediately from properties of the Smith form. We collect some of these properties in the next result, together with some brief remarks on their straightforward proofs.

Lemma 4.5 (Basic properties of Jordan characteristic).
Suppose $P(\lambda)$ and $Q(\lambda)$ are $m \times n$ matrix polynomials over an arbitrary field $\mathbb{F}$. Then

  (a) *$\mathcal{J}(\beta P, \mathbb{F}) = \mathcal{J}(P, \mathbb{F})$ for any nonzero $\beta \in \mathbb{F}$,*

  (b) *$\mathcal{J}(P^T, \mathbb{F}) = \mathcal{J}(P, \mathbb{F})$,*

  (c) *if $\mathbb{F} = \mathbb{C}$, then $\mathcal{J}(P^*, \lambda_0) = \mathcal{J}(\overline{P}, \lambda_0) = \mathcal{J}(P, \overline{\lambda}_0)$ for any $\lambda_0 \in \mathbb{C}_\infty$,*

  (d) *if P and Q are strictly equivalent, then $\mathcal{J}(P, \mathbb{F}) = \mathcal{J}(Q, \mathbb{F})$,*

  (e) *if P and Q are unimodularly equivalent, then $\mathcal{J}(P, \lambda_0) = \mathcal{J}(Q, \lambda_0)$ for all $\lambda_0 \in \mathbb{F}$ (but not necessarily for $\lambda_0 = \infty$).*

Further suppose that $P(\lambda)$ has grade k, and $\widetilde{P}(\lambda)$ is identical to $P(\lambda)$ except that $\widetilde{P}(\lambda)$ has grade $k + \ell$ for some $\ell \geq 1$. Then

(f) $\mathcal{J}(\widetilde{P}, \lambda_0) = \mathcal{J}(P, \lambda_0)$ *for all* $\lambda_0 \in \mathbb{F}$, *and* $\mathcal{J}(\widetilde{P}, \infty) = \mathcal{J}(P, \infty) + (\ell, \ell, \ldots, \ell)$.

Proof. Unimodularly equivalent polynomials certainly have the same Smith forms, so (e) follows; the fact that elementary divisors of unimodularly equivalent polynomials may differ at ∞ is discussed in [41]. Strictly equivalent polynomials are unimodularly equivalent, and so are their reversals, so (d) follows from (e). Part (a) is a special case of (d). For parts (b) and (c), note that if $D(\lambda)$ is the Smith form of $P(\lambda)$, then $D^T(\lambda)$ is the Smith form of $P^T(\lambda)$, $\overline{D}(\lambda)$ is the Smith form of $\overline{P}(\lambda)$, and similarly for the reversals, so (b) and (c) follow.

All that remains is to understand the effect of the choice of grade on the Jordan characteristic, as described in part (f). The first part of (f) is clear, since $\widetilde{P}$ has the same Smith form as P. To see the effect of grade on the elementary divisors at ∞, note that for $\widetilde{P}$ we must consider $\mathrm{rev}_{k+\ell}\widetilde{P}$, while for P the relevant reversal is $\mathrm{rev}_k P$. But it is easy to see that $\mathrm{rev}_{k+\ell}\widetilde{P}(\lambda) = \mathrm{rev}_{k+\ell}P(\lambda) = \lambda^\ell \mathrm{rev}_k P(\lambda)$. Thus the Smith forms of the reversals also differ by the factor λ^ℓ, and the second part of (f) now follows. $\square$

Remark 4.6. Clearly the Smith form of P determines the *finite* part of the Jordan characteristic, i.e., the part of the mapping $\mathcal{J}(P)$ restricted to $\mathbb{F} \subset \mathbb{F}_\infty$. When $\mathbb{F}$ is algebraically closed it is easy to see that the converse is also true; the Smith form of P can be uniquely reconstructed from the finite part of the Jordan characteristic. For fields $\mathbb{F}$ that are not algebraically closed, though, the Jordan characteristic $\mathcal{J}(P, \mathbb{F})$ over $\mathbb{F}$ will not always suffice to uniquely determine the Smith form of P over $\mathbb{F}$. Indeed, $\mathcal{J}(P, \lambda_0)$ can be the zero sequence for *all* $\lambda_0 \in \mathbb{F}$ when $\mathbb{F}$ is not algebraically closed, as the simple example of the *real* 1×1 polynomial $P(\lambda) = \lambda^2 + 1$ shows; thus the Jordan characteristic $\mathcal{J}(P, \mathbb{F})$ may not contain *any* information at all about the Smith form. However, because the Smith form of P over $\mathbb{F}$ is the *same* as the Smith form of P over the algebraic closure $\overline{\mathbb{F}}$, the Jordan characteristic $\mathcal{J}(P, \overline{\mathbb{F}})$ will uniquely determine the Smith form over $\mathbb{F}$, and not just over $\overline{\mathbb{F}}$.

There are many cases of interest where two matrix polynomials, e.g., a polynomial and any of its linearizations, have the same elementary divisors but different ranks, thus preventing the equality of their Jordan characteristics. In order to address this issue, it is convenient to introduce a truncated version of the Jordan characteristic in which all zeroes are discarded. For example, suppose P is a matrix polynomial whose Jordan characteristic at λ_0 is $\mathcal{J}(P, \lambda_0) = (0, 0, 0, 0, 1, 3, 4)$. Then the truncated Jordan characteristic of P at λ_0 will be just $(1, 3, 4)$. We define this formally as follows.

Definition 4.7 (Truncated Jordan characteristic).
Suppose P is a matrix polynomial over the field $\mathbb{F}$, and $\mathcal{J}(P, \lambda_0)$ is the Jordan characteristic of P at $\lambda_0 \in \mathbb{F}_\infty$. Then the truncated Jordan characteristic at λ_0, *denoted* $\widehat{\mathcal{J}}(P, \lambda_0)$, *is the nonzero subsequence of $\mathcal{J}(P, \lambda_0)$, i.e., the subsequence of $\mathcal{J}(P, \lambda_0)$ consisting of all of its nonzero entries. If $\mathcal{J}(P, \lambda_0)$ has only zero entries, then $\widehat{\mathcal{J}}(P, \lambda_0)$ is taken to be the empty sequence.*

Note that just as for the Jordan characteristic, it is natural to view the truncated Jordan characteristic over the field $\mathbb{F}$ as a mapping $\widehat{\mathcal{J}}(P)$ with domain $\mathbb{F}_\infty$. Clearly the values of this mapping are sequences whose lengths can vary from one element of $\mathbb{F}_\infty$ to another, and can be anything from zero up to $r = \mathrm{rank}(P)$.

Remark 4.8. The key feature of $\widehat{\mathcal{J}}(P, \lambda_0)$ is that it records *only* the information about the elementary divisors of P at λ_0, indeed, it is essentially the *Segre characteristic* of P at λ_0. Thus for *any* pair of matrix polynomials P and Q, regardless of their size, grade, or rank, the statement that P and Q have the same elementary divisors at λ_0 can now be

precisely captured by the equation $\widehat{\mathcal{J}}(P, \lambda_0) = \widehat{\mathcal{J}}(Q, \lambda_0)$. It also follows that the following two statements are equivalent:

1) $\mathcal{J}(P, \lambda_0) = \mathcal{J}(Q, \lambda_0)$

2) $\widehat{\mathcal{J}}(P, \lambda_0) = \widehat{\mathcal{J}}(Q, \lambda_0)$ *and* $\operatorname{rank}(P) = \operatorname{rank}(Q)$.

4.2 Tools for Calculating Jordan Characteristics

It is well known that for any $\lambda_0 \in \mathbb{F}$, any *scalar* polynomial $f(\lambda)$ over $\mathbb{F}$ can be uniquely expressed as $f(\lambda) = (\lambda - \lambda_0)^\alpha g(\lambda)$, where $\alpha \in \mathbb{N}$ and $g(\lambda_0) \neq 0$. (Note that $\alpha = 0 \in \mathbb{N}$ is certainly allowed here.) An analogous factorization for *matrix* polynomials is described in the following definition, which is well known for $\mathbb{F} = \mathbb{C}$, but also applicable for matrix polynomials over arbitrary fields.

Definition 4.9 (Local Smith representation)**.**
Let $P(\lambda)$ be an $m \times n$ matrix polynomial over a field $\mathbb{F}$, and $\lambda_0 \in \mathbb{F}$. Then a factorization

$$P(\lambda) = \big[E_{\lambda_0}(\lambda)\big]_{m \times m} \cdot \big[D_{\lambda_0}(\lambda)\big]_{m \times n} \cdot \big[F_{\lambda_0}(\lambda)\big]_{n \times n} \tag{4.3}$$

is called a local Smith representation *of P at $\lambda_0 \in \mathbb{F}$ if $E_{\lambda_0}(\lambda)$, $F_{\lambda_0}(\lambda)$ are matrix polynomials over $\mathbb{F}$ that are both invertible at $\lambda = \lambda_0$, and $D_{\lambda_0}(\lambda)$ is a diagonal $m \times n$ matrix polynomial of the form*

$$D_{\lambda_0}(\lambda) = \quad r \quad \begin{array}{c} \\ m-r \end{array} \overset{\displaystyle r \qquad\qquad n-r}{\left[\begin{array}{ccc|c} (\lambda - \lambda_0)^{\kappa_1} & & & \\ & \ddots & & \\ & & (\lambda - \lambda_0)^{\kappa_r} & \\ \hline & & & 0 \end{array}\right]}, \tag{4.4}$$

where $\big(\kappa_1, \kappa_2, \ldots, \kappa_r\big)$ are integer exponents satisfying $0 \leq \kappa_1 \leq \kappa_2 \leq \cdots \leq \kappa_r$.

As the name might suggest, local Smith representations of P are closely connected to the Smith form of P. Indeed, they can be viewed as weaker versions of the Smith form that focus attention on the behavior of $P(\lambda)$ near one constant λ_0 at a time. We will use them to develop tools for determining partial multiplicity sequences of $P(\lambda)$.

Because the polynomials $E_{\lambda_0}(\lambda)$, $F_{\lambda_0}(\lambda)$ in a local Smith representation do *not* need to be unimodular, only invertible at λ_0, tools based on local Smith representations can be much more flexible than the Smith form itself. The basic existence and uniqueness properties of local Smith representations, as well as their precise connection to the Smith form, are established in the following theorem.

Theorem 4.10 (Local Smith representations)**.**
Let $P(\lambda)$ be any $m \times n$ matrix polynomial, regular or singular, over an arbitrary field $\mathbb{F}$. Then for each $\lambda_0 \in \mathbb{F}$, there exists a local Smith representation (4.3) for P at λ_0. Moreover, in any two such representations (4.3) the diagonal factor $D_{\lambda_0}(\lambda)$ is the same, and so it is uniquely determined by P and λ_0. In particular, the number r of nonzero entries in $D_{\lambda_0}(\lambda)$ is always equal to $\operatorname{rank} P(\lambda)$, and the sequence of exponents $\big(\kappa_1, \kappa_2, \ldots, \kappa_r\big)$ is identical to the partial multiplicity sequence $\mathcal{J}(P, \lambda_0) = \big(\alpha_1, \alpha_2, \ldots, \alpha_r\big)$ determined by the Smith form of $P(\lambda)$.

Proof. The existence of a local Smith representation for $P(\lambda)$ at λ_0 follows easily from the Smith form itself, by doing a little extra factorization. Let $P(\lambda) = E(\lambda)D(\lambda)F(\lambda)$, where $E(\lambda)$, $F(\lambda)$ are unimodular and

$$D(\lambda) = \left[\begin{array}{ccc|c} d_1(\lambda) & & & \\ & \ddots & & \\ & & d_r(\lambda) & \\ \hline & & & 0 \end{array}\right]_{m \times n}$$

is the Smith form of $P(\lambda)$. Each $d_i(\lambda)$ factors uniquely as $d_i(\lambda) = (\lambda - \lambda_0)^{\alpha_i} \widetilde{d_i}(\lambda)$ where $\alpha_i \geq 0$ and $\widetilde{d_i}(\lambda_0) \neq 0$, so $D(\lambda)$ can be expressed as

$$D(\lambda) = \left[\begin{array}{ccc|c} (\lambda - \lambda_0)^{\alpha_1} & & & \\ & \ddots & & \\ & & (\lambda - \lambda_0)^{\alpha_r} & \\ \hline & & & 0 \end{array} \right]_{m \times n} \cdot \widetilde{D}(\lambda),$$

where $0 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_r$ and $\widetilde{D}(\lambda) = \mathrm{diag}\big[\widetilde{d_1}(\lambda), \widetilde{d_2}(\lambda), \ldots, \widetilde{d_r}(\lambda), 1, \ldots, 1\big]$ is an $n \times n$ diagonal matrix polynomial such that $\widetilde{D}(\lambda_0)$ is invertible. Thus

$$P(\lambda) = E_{\lambda_0}(\lambda) \cdot \left[\begin{array}{ccc|c} (\lambda - \lambda_0)^{\alpha_1} & & & \\ & \ddots & & \\ & & (\lambda - \lambda_0)^{\alpha_r} & \\ \hline & & & 0 \end{array} \right]_{m \times n} \cdot F_{\lambda_0}(\lambda), \qquad (4.5)$$

with $E_{\lambda_0}(\lambda) := E(\lambda)$ and $F_{\lambda_0}(\lambda) := \widetilde{D}(\lambda)F(\lambda)$ both invertible at λ_0, displays a local Smith representation for $P(\lambda)$ at λ_0.

Next we turn to the uniqueness part of the theorem. Suppose that we have two local Smith representations for $P(\lambda)$ at λ_0, given by $E_{\lambda_0}(\lambda)D_{\lambda_0}(\lambda)F_{\lambda_0}(\lambda)$ and $\widetilde{E}_{\lambda_0}(\lambda)\widetilde{D}_{\lambda_0}(\lambda)\widetilde{F}_{\lambda_0}(\lambda)$, i.e.,

$$E_{\lambda_0}(\lambda)D_{\lambda_0}(\lambda)F_{\lambda_0}(\lambda) = P(\lambda) = \widetilde{E}_{\lambda_0}(\lambda)\widetilde{D}_{\lambda_0}(\lambda)\widetilde{F}_{\lambda_0}(\lambda). \qquad (4.6)$$

Let $\big(\kappa_1, \kappa_2, \ldots, \kappa_r\big)$ and $\big(\widetilde{\kappa}_1, \widetilde{\kappa}_2, \ldots, \widetilde{\kappa}_{\widetilde{r}}\big)$ be the exponent sequences of $D_{\lambda_0}(\lambda)$ and $\widetilde{D}_{\lambda_0}(\lambda)$, respectively. We first show that the *lengths* r and $\widetilde{r}$ of these two sequences are the same. By hypothesis the *constant* matrices $E_{\lambda_0}(\lambda_0)$, $F_{\lambda_0}(\lambda_0)$, $\widetilde{E}_{\lambda_0}(\lambda_0)$, and $\widetilde{F}_{\lambda_0}(\lambda_0)$ are all invertible. Hence the corresponding square matrix *polynomials* $E_{\lambda_0}(\lambda)$, $F_{\lambda_0}(\lambda)$, $\widetilde{E}_{\lambda_0}(\lambda)$, and $\widetilde{F}_{\lambda_0}(\lambda)$ each have full rank when viewed as matrices over the field of rational functions $\mathbb{F}(\lambda)$. Thus (4.6) implies that $r = \mathrm{rank}\,D_{\lambda_0}(\lambda) = \mathrm{rank}\,P(\lambda) = \mathrm{rank}\,\widetilde{D}_{\lambda_0}(\lambda) = \widetilde{r}$.

To prove that the exponent sequences $\big(\kappa_1, \kappa_2, \ldots, \kappa_r\big)$ and $\big(\widetilde{\kappa}_1, \widetilde{\kappa}_2, \ldots, \widetilde{\kappa}_r\big)$ are identical, and hence that $D_{\lambda_0}(\lambda) \equiv \widetilde{D}_{\lambda_0}(\lambda)$, consider the partial sums $s_j := \kappa_1 + \kappa_2 + \cdots + \kappa_j$ and $\widetilde{s}_j := \widetilde{\kappa}_1 + \widetilde{\kappa}_2 + \cdots + \widetilde{\kappa}_j$ for $j = 1, \ldots, r$. If we can show that $s_j = \widetilde{s}_j$ for every $j = 1, \ldots, r$, then the desired identity of the two exponent sequences would follow.

Start by multiplying (4.6) on the left and right by the classical adjoint matrix polynomials $E_{\lambda_0}^{\#}(\lambda)$ and $F_{\lambda_0}^{\#}(\lambda)$, respectively, to obtain

$$\big[\det E_{\lambda_0}(\lambda)\,\det F_{\lambda_0}(\lambda)\big] \cdot D_{\lambda_0}(\lambda) = \widehat{E}_{\lambda_0}(\lambda)\widetilde{D}_{\lambda_0}(\lambda)\widehat{F}_{\lambda_0}(\lambda), \qquad (4.7)$$

where $\widehat{E}_{\lambda_0}(\lambda) = E_{\lambda_0}^{\#}(\lambda)\widetilde{E}_{\lambda_0}(\lambda)$ and $\widehat{F}_{\lambda_0}(\lambda) = \widetilde{F}_{\lambda_0}(\lambda)F_{\lambda_0}^{\#}(\lambda)$ are square matrix polynomials. For any fixed $1 \leq j \leq r$, taking the jth compound of both sides of (4.7) now yields

$$\big[\det E_{\lambda_0}(\lambda)\,\det F_{\lambda_0}(\lambda)\big]^j \cdot \mathcal{C}_j\big(D_{\lambda_0}(\lambda)\big) = \mathcal{C}_j\big(\widehat{E}_{\lambda_0}(\lambda)\big)\,\mathcal{C}_j\big(\widetilde{D}_{\lambda_0}(\lambda)\big)\,\mathcal{C}_j\big(\widehat{F}_{\lambda_0}(\lambda)\big), \qquad (4.8)$$

using the basic properties of compound matrices as described in Theorem 2.2. The jth compound of a diagonal matrix is also diagonal, and in particular

$$\mathcal{C}_j\big(D_{\lambda_0}(\lambda)\big) = \left[\begin{array}{cc} (\lambda - \lambda_0)^{s_j} & \\ & \ddots \end{array} \right] \quad \text{and} \quad \mathcal{C}_j\big(\widetilde{D}_{\lambda_0}(\lambda)\big) = \left[\begin{array}{cc} (\lambda - \lambda_0)^{\widetilde{s}_j} & \\ & \ddots \end{array} \right].$$

Note that *every* nonzero diagonal entry of $\mathcal{C}_j\big(\widetilde{D}_{\lambda_0}(\lambda)\big)$ is of the form $(\lambda - \lambda_0)^{\ell}$ with $\ell \geq \widetilde{s}_j$, so $\mathcal{C}_j\big(\widetilde{D}_{\lambda_0}(\lambda)\big)$, and hence also the right-hand side of (4.8), may be factored as

$$(\lambda - \lambda_0)^{\widetilde{s}_j}Q(\lambda),$$

17

where $Q(\lambda)$ is a matrix *polynomial*. Hence, every entry of the left-hand side of (4.8) must be divisible by $(\lambda - \lambda_0)^{\widetilde{s}_j}$, in particular the $(1,1)$-entry $\big[\det E_{\lambda_0}(\lambda) \, \det F_{\lambda_0}(\lambda) \big]^j (\lambda - \lambda_0)^{s_j}$. But the invertibility of $E_{\lambda_0}(\lambda)$ and $F_{\lambda_0}(\lambda)$ at λ_0 means that neither $\det E_{\lambda_0}(\lambda)$ nor $\det F_{\lambda_0}(\lambda)$ are divisible by $(\lambda - \lambda_0)$. Thus $(\lambda - \lambda_0)^{\widetilde{s}_j}$ divides $(\lambda - \lambda_0)^{s_j}$, and hence $\widetilde{s}_j \leq s_j$ for each $j = 1, \ldots, r$. Interchanging the roles of the left and right sides of (4.6), we see by the same argument, mutatis mutandis, that $s_j \leq \widetilde{s}_j$ for $j = 1, \ldots, r$. Thus $s_j = \widetilde{s}_j$ for each $j = 1, \ldots, r$, and therefore the exponent sequences $(\kappa_1, \kappa_2, \ldots, \kappa_r)$ and $(\widetilde{\kappa}_1, \widetilde{\kappa}_2, \ldots, \widetilde{\kappa}_r)$ are identical.

Finally, the existence of the particular local Smith representation (4.5) generated from the Smith form shows that the exponent sequence $(\kappa_1, \kappa_2, \ldots, \kappa_r)$ from every local Smith representation is always the same as the partial multiplicity sequence $\mathcal{J}(P, \lambda_0) = (\alpha_1, \alpha_2, \ldots, \alpha_r)$. $\square$

Definition 4.11 (Local Smith form).
The uniquely determined diagonal factor $D_{\lambda_0}(\lambda)$ in (4.3) will be referred to as the local Smith form *of $P(\lambda)$ at λ_0.*

Remark 4.12. The proof of Theorem 4.10 presented here has been designed to be valid for *general* matrix polynomials over *arbitrary* fields, and also to take advantage of the properties of compound matrices. With some modifications, an earlier proof [30, Theorem S1.10] for *regular* matrix polynomials over $\mathbb{C}$ can be made valid for general matrix polynomials over arbitrary fields. Further extensions of the concept of local Smith form (at least for $\mathbb{F} = \mathbb{C}$) can also be found in the literature. For example, the conditions in Definition 4.9 can be relaxed as in [40] to allow $E_{\lambda_0}(\lambda)$ and $F_{\lambda_0}(\lambda)$ to be any matrix functions that are invertible at λ_0 and analytic in a neighborhood of λ_0 (e.g., rational matrices whose determinants have no zeroes or poles at λ_0). This extension was used in [3] to study linearizations.

As a consequence of the uniqueness of the local Smith form and Theorem 4.10, we obtain the following useful tool for determining partial multiplicity sequences.

Lemma 4.13. *Suppose $P(\lambda)$ and $Q(\lambda)$ are $m \times n$ matrix polynomials, and let $\lambda_0 \in \mathbb{F}$. If*

$$G_{\lambda_0}(\lambda) \, P(\lambda) \, H_{\lambda_0}(\lambda) \; = \; K_{\lambda_0}(\lambda) \, Q(\lambda) \, M_{\lambda_0}(\lambda),$$

where $G_{\lambda_0}(\lambda), H_{\lambda_0}(\lambda), K_{\lambda_0}(\lambda), M_{\lambda_0}(\lambda)$ are matrix polynomials invertible at λ_0, then the partial multiplicity sequences $\mathcal{J}(P, \lambda_0)$ and $\mathcal{J}(Q, \lambda_0)$ are identical.

Proof. Define $R(\lambda) := G_{\lambda_0}(\lambda)P(\lambda)H_{\lambda_0}(\lambda)$, and let $P(\lambda) := E_{\lambda_0}(\lambda)D_{\lambda_0}(\lambda)F_{\lambda_0}(\lambda)$ be a local Smith representation for $P(\lambda)$ as in Theorem 4.10. Then

$$R(\lambda) \; = \; G_{\lambda_0}(\lambda)P(\lambda)H_{\lambda_0}(\lambda) \; = \; \big[G_{\lambda_0}(\lambda)E_{\lambda_0}(\lambda) \big] D_{\lambda_0}(\lambda) \big[F_{\lambda_0}(\lambda)H_{\lambda_0}(\lambda) \big]$$

displays a local Smith representation for $R(\lambda)$ at λ_0. Since the local Smith form at λ_0 is unique, we see that $\mathcal{J}(P, \lambda_0)$ must be identical to $\mathcal{J}(R, \lambda_0)$. The same argument applies to $\mathcal{J}(Q, \lambda_0)$ and $\mathcal{J}(R, \lambda_0)$, which implies the desired result. $\square$

Remark 4.14. An extension of Lemma 4.13 in which the matrix polynomials $G_{\lambda_0}(\lambda), H_{\lambda_0}(\lambda), K_{\lambda_0}(\lambda), M_{\lambda_0}(\lambda)$ are replaced by *rational* matrices invertible at λ_0 follows easily as a corollary of Lemma 4.13. Although this extension does not in fact provide any greater generality than Lemma 4.13, it may certainly at times be more convenient to use.

4.3 Jordan Characteristic and Smith Form of $\mathrm{rev}_k P$

As a first illustration of the use of Lemma 4.13, we show how the Jordan characteristic of a matrix polynomial $P(\lambda)$ is related to that of its reversal polynomial $\mathrm{rev}_k P$. Recall from Proposition 3.29 that $\mathrm{rank}(\mathrm{rev}_k P) = \mathrm{rank}\, P = r$, so all partial multiplicity sequences of both P and $\mathrm{rev}_k P$ have length r, and hence can be compared to each other.

Theorem 4.15 (Partial multiplicity sequences of $\mathrm{rev}_k P$).
Suppose $P(\lambda)$ is an $m \times n$ matrix polynomial with $\mathrm{grade}(P) = k$ and $\mathrm{rank}(P) = r$. Then

$$\mathcal{J}(\mathrm{rev}_k P, \lambda_0) \equiv \mathcal{J}(P, 1/\lambda_0) \quad \text{for all} \quad \lambda_0 \in \mathbb{F}_\infty \,.$$

(Here 0 and ∞ are also included as a reciprocal pair.) Equivalently, we have that the following diagram commutes, where m_R is the reciprocal map $\lambda_0 \mapsto 1/\lambda_0$ on $\mathbb{F}_\infty$.

$$
\begin{array}{ccc}
\mathbb{F}_\infty & \xrightarrow{\ \mathcal{J}(\mathrm{rev}_k P)\ } & \mathbb{N}^r_\leq \\
{\scriptstyle \mathsf{m}_R}\downarrow & \nearrow {\scriptstyle \mathcal{J}(P)} & \\
\mathbb{F}_\infty & &
\end{array}
$$

Proof. First observe that $\mathcal{J}(P, \infty)$ is *by definition* the same as $\mathcal{J}(\mathrm{rev}_k P, 0)$. Similarly we see that $\mathcal{J}(\mathrm{rev}_k P, \infty) = \mathcal{J}\big(\mathrm{rev}_k(\mathrm{rev}_k P), 0\big) = \mathcal{J}(P, 0)$, so to complete the proof we from now on restrict attention to nonzero finite $\lambda_0 \in \mathbb{F}$. Set $\mu_0 = \frac{1}{\lambda_0}$, and let

$$P(\lambda) = E_{\mu_0}(\lambda) D_{\mu_0}(\lambda) F_{\mu_0}(\lambda) \tag{4.9}$$

be a local Smith representation of $P(\lambda)$ at μ_0, where

$$D_{\mu_0}(\lambda) \;=\; \mathrm{diag}\big[\, (\lambda - \mu_0)^{\alpha_1}, (\lambda - \mu_0)^{\alpha_2}, \ldots, (\lambda - \mu_0)^{\alpha_r}, 0, \ldots, 0 \,\big]\,,$$

as in (4.4). Now define $\ell := \deg\big(E_{\mu_0}(\lambda)\big) + \deg\big(D_{\mu_0}(\lambda)\big) + \deg\big(F_{\mu_0}(\lambda)\big)$, let $s := \max\{k, \ell\}$, and take the s-reversal of both sides of (4.9) to obtain

$$\lambda^{s-k} \, \mathrm{rev}_k P(\lambda) \;=\; \lambda^{s-\ell} \, \mathrm{rev}\, E_{\mu_0}(\lambda) \, \mathrm{rev}\, D_{\mu_0}(\lambda) \, \mathrm{rev}\, F_{\mu_0}(\lambda)\,,$$

where each reversal on the right-hand side is taken with respect to degree. Since λ^{s-k}, $\lambda^{s-\ell}$, $\mathrm{rev}\, E_{\mu_0}(\lambda)$ and $\mathrm{rev}\, F_{\mu_0}(\lambda)$ are all matrix polynomials that are invertible at $\lambda_0 = \frac{1}{\mu_0} \neq 0$, by Lemma 4.13 we see that $\mathcal{J}(\mathrm{rev}_k P, \lambda_0)$ and $\mathcal{J}(\mathrm{rev}\, D_{\mu_0}, \lambda_0)$ are identical. But

$$\mathrm{rev}\, D_{\mu_0}(\lambda) \;=\; \lambda^{\alpha_r} D_{\mu_0}(1/\lambda)$$

$$= \; \lambda^{\alpha_r}
\left[
\begin{array}{ccc|c}
\left(\frac{-\mu_0}{\lambda}\right)^{\alpha_1} & & & \\
& \ddots & & \\
& & \left(\frac{-\mu_0}{\lambda}\right)^{\alpha_r} & \\
\hline
& & & I_{m-r}
\end{array}
\right]_{m \times m}
\left[
\begin{array}{ccc|c}
\left(\lambda - \lambda_0\right)^{\alpha_1} & & & \\
& \ddots & & \\
& & \left(\lambda - \lambda_0\right)^{\alpha_r} & \\
\hline
& & & 0
\end{array}
\right]_{m \times n}$$

displays a local Smith representation for $\mathrm{rev}\, D_{\mu_0}$ at λ_0, showing that $\mathcal{J}(\mathrm{rev}\, D_{\mu_0}, \lambda_0) = (\alpha_1, \alpha_2, \ldots, \alpha_r)$ is exactly the same as $\mathcal{J}(P, \mu_0) = \mathcal{J}(P, 1/\lambda_0)$, thus completing the proof. $\square$

Remark 4.16. Theorem 4.15 is well-known for special fields and was stated (without proof) for polynomials over $\mathbb{C}$ in [19].

For polynomials over non-algebraically closed fields $\mathbb{F}$, the result of Theorem 4.15 does not completely capture the effect of reversal on the Smith form. We characterize the general relationship between the Smith forms of P and $\mathrm{rev}\, P$ in Theorem 4.17 below, so the reader may see the full picture for the reversal operation together in the same place. However, no proof will be given here, since this result is just a special case of Theorem 5.7. Note that nothing in Section 5 needs Theorem 4.17 as a prerequisite, so there is no logical circularity.

Theorem 4.17 (Smith forms of P and $\mathrm{rev}_k\, P$)**.**
Suppose $P(\lambda)$ is an $m \times n$ matrix polynomial over a field $\mathbb{F}$ with $\mathrm{grade}(P) = k$ and $\mathrm{rank}(P) = r$. Let

$$
D(\lambda) \;=\; \left[
\begin{array}{ccc|c}
\lambda^{\alpha_1} p_1(\lambda) & & & \\
 & \ddots & & \\
 & & \lambda^{\alpha_r} p_r(\lambda) & \\
\hline
 & & & 0
\end{array}
\right]_{m \times n}
\qquad with \quad p_j(0) \neq 0 \ \ for \ \ 1 \leq j \leq r
$$

be the Smith form of $P(\lambda)$, and suppose $\mathcal{J}(P, \infty) = \big(\beta_1, \beta_2, \ldots, \beta_r\big)$. Then the Smith form of $\mathrm{rev}_k\, P$ is

$$
\widetilde{D}(\lambda) \;=\; \left[
\begin{array}{ccc|c}
\gamma_1 \lambda^{\beta_1} \mathrm{rev}\, p_1(\lambda) & & & \\
 & \ddots & & \\
 & & \gamma_r \lambda^{\beta_r} \mathrm{rev}\, p_r(\lambda) & \\
\hline
 & & & 0
\end{array}
\right]_{m \times n} ,
$$

where the constants γ_j for $1 \leq j \leq r$ are chosen so that each $\gamma_j\, \mathrm{rev}\, p_j(\lambda)$ is monic. (Note that each reversal $\mathrm{rev}\, p_j(\lambda)$ in $\widetilde{D}(\lambda)$ is taken with respect to $\deg p_j$.)

5 Möbius and Jordan and Smith

We now consider the elementary divisors of a general matrix polynomial P, and their relationship to the elementary divisors of any of the Möbius transforms of P. This relationship can be conveniently and concisely expressed by comparing the Jordan characteristics and Smith forms of P and $\mathbf{M}_A(P)$; these comparisons constitute the two main results of this section, Theorems 5.3 and 5.7.

Alternative approaches to this elementary divisor question are taken in [5], [6], and [66], with results analogous to those in this section. In [66], matrix polynomials and their Smith forms are treated in homogeneous form, but the only choice for grade considered is grade $P = \deg P$. The analysis is extended in [5] and [6] to include not just matrices with entries in a ring of polynomials, but to rational matrices, and even matrices with entries from an arbitrary local ring. See also Remark 5.5 for an extension to more general rational transforms.

5.1 Möbius and Jordan Characteristic

Before proceeding to the main result of this section, we present two preliminary lemmas containing some observations about Möbius functions that will be needed later.

Lemma 5.1. *Let $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in GL(2, \mathbb{F})$ with corresponding Möbius function $\mathsf{m}_A(\lambda) = \frac{a\lambda+b}{c\lambda+d}$, and $R = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ as in Lemma 3.3.*

(a) *If $\mathsf{m}_A(\infty) = \infty$, then $\mathsf{m}_{RAR}(0) = 0$, in particular $\mathsf{m}_{RAR}(0)$ is finite.*

(b) *If $\mathsf{m}_A(\infty) = \lambda_0$ is finite and nonzero, then $\mathsf{m}_{RAR}(0) = \frac{1}{\lambda_0}$ is finite.*

(c) *If $\mathsf{m}_A(\infty) = 0$, then m_A can be factored as the composition of two Möbius functions $\mathsf{m}_{A_2}\mathsf{m}_{A_1} = \mathsf{m}_{A_2 A_1} = \mathsf{m}_A$ such that $\mathsf{m}_{A_1}(\infty) = 1$ and $\mathsf{m}_{A_2}(1) = 0$.*

Proof. Parts (a) and (b) are special cases of Lemma 3.3. Since

$$
\mathsf{m}_{RAR}(0) \;=\; \mathsf{m}_{RAR}\left(\frac{1}{\infty}\right) \;=\; \frac{1}{\mathsf{m}_A(\infty)} \,,
$$

for part (a) we have $\mathsf{m}_{RAR}(0) = \frac{1}{\infty} = 0$, and for part (b) we have $\mathsf{m}_{RAR}(0) = \frac{1}{\lambda_0}$.

For part (c), it suffices to find an appropriate factorization $A_2 A_1$ for A. First observe that $\mathsf{m}_A(\infty) = 0$ implies that $A\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ c \end{bmatrix}$ for some nonzero c; hence $A = \begin{bmatrix} 0 & b \\ c & d \end{bmatrix}$ with $b \neq 0$ and $c \neq 0$, since A is nonsingular. Then

$$A_2 A_1 = \begin{bmatrix} -b & b \\ c-d & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & b \\ c & d \end{bmatrix} = A$$

is one such factorization, since $A_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ implies that $\mathsf{m}_{A_1}(\infty) = 1$, and $A_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ c \end{bmatrix}$ with $c \neq 0$ implies that $\mathsf{m}_{A_2}(1) = 0$. $\qquad\square$

Lemma 5.2. *Suppose* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL(2, \mathbb{F})$, *with corresponding Möbius function* $\mathsf{m}_A(\lambda) = \frac{a\lambda+b}{c\lambda+d}$. *If* μ_0 *and* $\mathsf{m}_A(\mu_0)$ *are both finite elements of* $\mathbb{F}_\infty$, *then we have the following identity in the field of rational functions* $\mathbb{F}(\mu)$:

$$(c\mu + d)\big[\mathsf{m}_A(\mu) - \mathsf{m}_A(\mu_0)\big] = \frac{\det A}{(c\mu_0 + d)}(\mu - \mu_0).$$

Proof. Since $\mathsf{m}_A(\mu_0)$ is finite, the quantity $c\mu_0 + d$ is nonzero. Thus from the simple calculation

$$\begin{aligned}
(c\mu_0 + d)(c\mu + d)\big[\mathsf{m}_A(\mu) - \mathsf{m}_A(\mu_0)\big] &= \big[(a\mu + b)(c\mu_0 + d) - (a\mu_0 + b)(c\mu + d)\big] \\
&= \big[(ad - bc)\mu - (ad - bc)\mu_0\big] \\
&= (\det A)(\mu - \mu_0),
\end{aligned}$$

the result follows immediately. $\qquad\square$

Theorem 5.3 (Partial multiplicity sequences of Möbius transforms)**.**
Let $P(\lambda)$ *be an* $m \times n$ *matrix polynomial over a field* $\mathbb{F}$ *with* $\mathrm{grade}(P) = k$ *and* $\mathrm{rank}(P) = r$, *and let* $A \in GL(2, \mathbb{F})$ *with associated Möbius transformation* $\mathbf{M}_A$ *and Möbius function* m_A. *Then for any* $\mu_0 \in \mathbb{F}_\infty$,

$$\mathcal{J}\big(\mathbf{M}_A(P), \mu_0\big) \equiv \mathcal{J}\big(P, \mathsf{m}_A(\mu_0)\big). \tag{5.1}$$

Equivalently, we may instead write

$$\mathcal{J}\big(\mathbf{M}_A(P), \mathsf{m}_{A^{-1}}(\mu_0)\big) \equiv \mathcal{J}\big(P, \mu_0\big), \tag{5.2}$$

or assert that the following diagram commutes.

Proof. The proof proceeds in five cases, depending on various combinations of the numbers μ_0 and $\lambda_0 := \mathsf{m}_A(\mu_0)$ being finite, infinite, or nonzero. Throughout the proof we have $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

<u>Case 1</u> $[\mu_0$ and $\lambda_0 = \mathsf{m}_A(\mu_0)$ are both finite$]$:   First observe that $\mathsf{m}_A(\mu_0)$ being finite means that $c\mu_0 + d \neq 0$, a fact that will be used implicitly throughout the argument. Now let

$$P(\lambda) = E_{\lambda_0}(\lambda) D_{\lambda_0}(\lambda) F_{\lambda_0}(\lambda) \tag{5.3}$$

be a local Smith representation for $P(\lambda)$ at λ_0, where $D_{\lambda_0}(\lambda)$ is as in (4.4); hence by Theorem 4.10 we have $(\kappa_1, \kappa_2, \ldots, \kappa_r) = \mathcal{J}(P, \lambda_0) = (\alpha_1, \alpha_2, \ldots, \alpha_r)$ is the partial multiplicity

sequence of P at λ_0, and $\deg D_{\lambda_0}(\lambda) = \alpha_r$. Now define $\ell := \deg\big(E_{\lambda_0}(\lambda)\big) + \deg\big(D_{\lambda_0}(\lambda)\big) + \deg\big(F_{\lambda_0}(\lambda)\big)$, and let $s := \max\{k, \ell\}$. Then by substituting $\lambda = \mathsf{m}_A(\mu)$ in (5.3) and pre-multiplying by $(c\mu + d)^s$ we obtain

$$(c\mu + d)^s P\big(\mathsf{m}_A(\mu)\big) \;=\; (c\mu + d)^s E_{\lambda_0}\big(\mathsf{m}_A(\mu)\big) D_{\lambda_0}\big(\mathsf{m}_A(\mu)\big) F_{\lambda_0}\big(\mathsf{m}_A(\mu)\big),$$

or equivalently

$$(c\mu + d)^{s-k}\big[\, \mathbf{M}_A(P)(\mu)\,\big]$$
$$= \Big[(c\mu + d)^{s-\ell}\, \mathbf{M}_A\big(E_{\lambda_0}\big)(\mu)\Big] \cdot \Big[(c\mu + d)^{\alpha_r} D_{\lambda_0}\big(\mathsf{m}_A(\mu)\big)\Big] \cdot \Big[\mathbf{M}_A\big(F_{\lambda_0}\big)(\mu)\Big],$$

where the Möbius transforms $\mathbf{M}_A\big(E_{\lambda_0}\big)$ and $\mathbf{M}_A\big(F_{\lambda_0}\big)$ are taken with respect to degree in each case. Since $(c\mu + d)^{s-k}$, $(c\mu + d)^{s-\ell}$, $\mathbf{M}_A\big(E_{\lambda_0}\big)(\mu)$, and $\mathbf{M}_A\big(F_{\lambda_0}\big)(\mu)$ are all matrix polynomials that are invertible at $\mu = \mu_0$, we see by Lemma 4.13 that

$$\mathcal{J}\big(\mathbf{M}_A(P), \mu_0\big) \;\equiv\; \mathcal{J}\Big((c\mu + d)^{\alpha_r} D_{\lambda_0}\big(\mathsf{m}_A(\mu)\big), \mu_0\Big).$$

We can now simplify $(c\mu + d)^{\alpha_r} D_{\lambda_0}\big(\mathsf{m}_A(\mu)\big)$ even further using Lemma 5.2 and obtain

$$(c\mu + d)^{\alpha_r} D_{\lambda_0}\big(\mathsf{m}_A(\mu)\big) \;=\; (c\mu + d)^{\alpha_r}\left[\begin{array}{cccc|c} \big(\mathsf{m}_A(\mu) - \lambda_0\big)^{\alpha_1} & & & & \\ & \ddots & & & \\ & & \big(\mathsf{m}_A(\mu) - \lambda_0\big)^{\alpha_r} & & \\ \hline & & & & 0 \end{array}\right]_{m \times n}$$

$$= (c\mu + d)^{\alpha_r}\left[\begin{array}{cccc|c} \big(\mathsf{m}_A(\mu) - \mathsf{m}_A(\mu_0)\big)^{\alpha_1} & & & & \\ & \ddots & & & \\ & & \big(\mathsf{m}_A(\mu) - \mathsf{m}_A(\mu_0)\big)^{\alpha_r} & & \\ \hline & & & & 0 \end{array}\right]_{m \times n}$$

$$= G(\mu)\left[\begin{array}{cccc|c} \big(\mu - \mu_0\big)^{\alpha_1} & & & & \\ & \ddots & & & \\ & & \big(\mu - \mu_0\big)^{\alpha_r} & & \\ \hline & & & & 0 \end{array}\right]_{m \times n}, \qquad (5.4)$$

where $G(\mu)$ is the diagonal matrix polynomial

$$G(\mu) = \left[\begin{array}{ccccc|c} \beta^{\alpha_1}(c\mu + d)^{\alpha_r - \alpha_1} & & & & & \\ & \beta^{\alpha_2}(c\mu + d)^{\alpha_r - \alpha_2} & & & & \\ & & \ddots & & & \\ & & & \beta^{\alpha_r}(c\mu + d)^{\alpha_r - \alpha_r} & & \\ \hline & & & & & I_{m-r} \end{array}\right]_{m \times m} \quad \text{with} \quad \beta = \frac{\det A}{(c\mu_0 + d)}.$$

Since $G(\mu)$ is invertible at $\mu = \mu_0$, we can now read off from the local Smith representation (5.4) that

$$\mathcal{J}\Big((c\mu + d)^{\alpha_r} D_{\lambda_0}\big(\mathsf{m}_A(\mu)\big), \mu_0\Big) \;=\; (\alpha_1, \alpha_2, \ldots, \alpha_r)$$
$$= \mathcal{J}(P, \lambda_0) \;=\; \mathcal{J}\big(P, \mathsf{m}_A(\mu_0)\big),$$

and the proof for Case 1 is complete.

Case 2 [μ_0 and $\lambda_0 = \mathsf{m}_A(\mu_0)$ are both ∞]:

Using the fact that R as in Example 3.7, Definition 4.2, and Lemma 5.1(a) satisfies $R^2 = I$, we have

$$
\begin{aligned}
\mathcal{J}\big(\mathbf{M}_A(P), \infty\big) \;=\; \mathcal{J}\Big(\mathbf{M}_R\big(\mathbf{M}_A(P)\big), 0\Big) \;&=\; \mathcal{J}\big(\mathbf{M}_{AR}(P), 0\big) \\
&=\; \mathcal{J}\Big(\mathbf{M}_{RAR}\big(\mathbf{M}_R(P)\big), 0\Big) \\
&=\; \mathcal{J}\big(\mathbf{M}_R(P), \mathsf{m}_{RAR}(0)\big) \\
&=\; \mathcal{J}\big(\mathbf{M}_R(P), 0\big) \;=\; \mathcal{J}\big(P, \infty\big) \;=\; \mathcal{J}\big(P, \mathsf{m}_A(\mu_0)\big).
\end{aligned}
$$

The first equality is by Definition 4.2, the second and third by Theorem 3.18(b) together with $R^2 = I$, the fourth and fifth by Case 1 together with Lemma 5.1(a), and the last equality by Definition 4.2.

Case 3 [$\mu_0 = \infty$, but $\lambda_0 = \mathsf{m}_A(\mu_0)$ is finite and *nonzero*]:

$$
\begin{aligned}
\mathcal{J}\big(\mathbf{M}_A(P), \infty\big) \;=\; \mathcal{J}\Big(\mathbf{M}_{RAR}\big(\mathbf{M}_R(P)\big), 0\Big) \;&=\; \mathcal{J}\big(\mathbf{M}_R(P), \mathsf{m}_{RAR}(0)\big) \\
&=\; \mathcal{J}\big(\mathbf{M}_R(P), 1/\lambda_0\big) \\
&=\; \mathcal{J}\big(P, \lambda_0\big) \;=\; \mathcal{J}\big(P, \mathsf{m}_A(\mu_0)\big).
\end{aligned}
$$

The first equality condenses the first three equalities of Case 2, the second and third are by Lemma 5.1(b) together with Case 1, the fourth equality is by Theorem 4.15, and the last equality just uses the definition of λ_0 for this case.

Case 4 [$\mu_0 = \infty$ and $\lambda_0 = \mathsf{m}_A(\mu_0) = 0$]: This case uses Lemma 5.1(c), which guarantees that any m_A with the property that $\mathsf{m}_A(\infty) = 0$ can always be factored as a composition of Möbius functions $\mathsf{m}_{A_2}\mathsf{m}_{A_1} = \mathsf{m}_A$ such that $\mathsf{m}_{A_1}(\infty) = 1$ and $\mathsf{m}_{A_2}(1) = 0$. Using such a factorization we can now prove Case 4 as a consequence of Case 3 and Case 1.

$$
\begin{aligned}
\mathcal{J}\big(\mathbf{M}_A(P), \mu_0\big) \;=\; \mathcal{J}\big(\mathbf{M}_{A_2 A_1}(P), \infty\big) \;&=\; \mathcal{J}\Big(\mathbf{M}_{A_1}\big(\mathbf{M}_{A_2}(P)\big), \infty\Big) \\
&=\; \mathcal{J}\big(\mathbf{M}_{A_2}(P), \mathsf{m}_{A_1}(\infty)\big) \\
&=\; \mathcal{J}\big(\mathbf{M}_{A_2}(P), 1\big) \\
&=\; \mathcal{J}\big(P, \mathsf{m}_{A_2}(1)\big) \\
&=\; \mathcal{J}\big(P, 0\big) \;=\; \mathcal{J}\big(P, \mathsf{m}_A(\mu_0)\big).
\end{aligned}
$$

The first equality invokes the factorization, the second is by Theorem 3.18(b), the third equality is by Case 3, the fifth equality is by Case 1, and the last equality uses the definition of λ_0 for this case.

Case 5 [μ_0 is finite, and $\lambda_0 = \mathsf{m}_A(\mu_0) = \infty$]: For this final case we first need to observe that $\mathsf{m}_A(\mu_0) = \infty$ implies that $\mathsf{m}_{A^{-1}}(\infty) = \mu_0$.

$$
\begin{aligned}
\mathcal{J}\big(\mathbf{M}_A(P), \mu_0\big) \;&=\; \mathcal{J}\big(\mathbf{M}_A(P), \mathsf{m}_{A^{-1}}(\infty)\big) \\
&=\; \mathcal{J}\Big(\mathbf{M}_{A^{-1}}\big(\mathbf{M}_A(P)\big), \infty\Big) \;=\; \mathcal{J}\big(P, \infty\big) \;=\; \mathcal{J}\big(P, \mathsf{m}_A(\mu_0)\big).
\end{aligned}
$$

The first equality invokes the observation, the second uses Case 3 and 4, the third follows from Theorem 3.18(c), and the last equality uses the definition of λ_0 for this case. $\qquad\square$

Theorem 5.3 provides a beautiful and substantial generalization of Theorem 4.15, since the reversal operator $\mathrm{rev}_k P$ is just the particular Möbius transformation induced by the matrix R, as described in Example 3.7.

Remark 5.4. It should be noted that the relationship proved in Theorem 5.3 also holds for truncated Jordan characteristic, i.e.,

$$\widehat{\mathcal{J}}\big(\mathbf{M}_A(P), \lambda_0\big) \;=\; \widehat{\mathcal{J}}\big(P, \mathsf{m}_A(\lambda_0)\big) \quad \text{for every} \quad \lambda_0 \in \mathbb{F}_\infty \,,$$

for any matrix polynomial P and any Möbius transformation $\mathbf{M}_A$. This follows immediately by applying Remark 4.8 to the corresponding result in Theorem 5.3 for the "ordinary" Jordan characteristic.

Remark 5.5. In view of the results in this section, we suggested that the Jordan characteristic may behave in a manner analogous to (5.1) under more general *rational transforms* of matrix polynomials. This topic has been considered in work by V. Noferini [53].

5.2   Möbius and the Smith Form

The results of Section 5.1 now enable us to find explicit formulas for the Smith form of any Möbius transform of a matrix polynomial P in terms of the Smith form of P itself. To derive these we will make use of the following lemma.

Lemma 5.6 (Möbius transforms of scalar polynomials)**.**
Suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL(2, \mathbb{F})$, with corresponding Möbius function $\mathsf{m}_A(\lambda) = \frac{a\lambda + b}{c\lambda + d}$. Consider any scalar polynomial of the form $p(\lambda) = (\lambda - \lambda_1)^{\alpha_1}(\lambda - \lambda_2)^{\alpha_2} \cdots (\lambda - \lambda_k)^{\alpha_k}$ where $\lambda_j \in \mathbb{F}$ are distinct *finite numbers. Let $r = \alpha_1 + \alpha_2 + \cdots + \alpha_k$, and define μ_j to be the unique distinct elements of $\mathbb{F}_\infty$ so that $\lambda_j = \mathsf{m}_A(\mu_j)$ for $j = 1, \ldots, k$.*

(a) *If $\lambda_j \neq \mathsf{m}_A(\infty)$ for $j = 1, \ldots, k$, then*

$$\big[\mathbf{M}_A(p)\big](\mu) \;=\; \gamma(\mu - \mu_1)^{\alpha_1}(\mu - \mu_2)^{\alpha_2} \cdots (\mu - \mu_k)^{\alpha_k} \,,$$

where $\mathbf{M}_A$ is taken with respect to degree, and

$$\gamma \;=\; \frac{(\det A)^r}{(c\mu_1 + d)^{\alpha_1}(c\mu_2 + d)^{\alpha_2} \cdots (c\mu_k + d)^{\alpha_k}} \,,$$

is a finite nonzero constant. (Note that all terms in the denominator of γ are nonzero since each $\mathsf{m}_A(\mu_j) = \lambda_j$ is finite.)

(b) *If $\lambda_1 = \mathsf{m}_A(\infty)$, and so $\lambda_j \neq \mathsf{m}_A(\infty)$ for all $j = 2, \ldots, k$, then*

$$\big[\mathbf{M}_A(p)\big](\mu) \;=\; \widetilde{\gamma}\, (\mu - \mu_2)^{\alpha_2} \cdots (\mu - \mu_k)^{\alpha_k} \,,$$

where $\mathbf{M}_A$ is taken with respect to degree, and

$$\widetilde{\gamma} \;=\; \frac{(\det A)^r}{(-c)^{\alpha_1}(c\mu_2 + d)^{\alpha_2} \cdots (c\mu_k + d)^{\alpha_k}} \,,$$

is a nonzero constant. (Note that all terms in the denominator of γ are nonzero since each $\mathsf{m}_A(\mu_j) = \lambda_j$ is finite. Also $\mathsf{m}_A(\infty) = \lambda_1$ being finite implies that $c \neq 0$.)

Observe that in this case $\deg \mathbf{M}_A(p)$ is strictly less than $\deg p$, since the term $(\lambda - \lambda_1)^{\alpha_1}$ is effectively swallowed up by the Möbius transform $\mathbf{M}_A$.

Proof. (a) First consider the case where $p(\lambda)$ consists of a single linear factor $(\lambda - \lambda_j)$. Then by the definition of $\mathbf{M}_A$ and Lemma 5.2 we have

$$\mathbf{M}_A(\lambda - \lambda_j) \;=\; (c\mu + d)\big[\mathsf{m}_A(\mu) - \lambda_j\big] \;=\; (c\mu + d)\big[\mathsf{m}_A(\mu) - \mathsf{m}_A(\mu_j)\big] \;=\; \frac{\det A}{(c\mu_j + d)}\,(\mu - \mu_j).$$

The result of part (a) now follows by applying the multiplicative property of Möbius transformations as described in Corollary 3.24(a).

(b) First observe that λ_1 being finite with $\lambda_1 = \mathsf{m}_A(\infty)$ means that $\lambda_1 = a/c$ with $c \neq 0$. Then we have

$$\mathbf{M}_A(\lambda - \lambda_1) = (c\mu + d)\left[\frac{a\mu + b}{c\mu + d} - \frac{a}{c}\right] = (a\mu + b) - \frac{a}{c}(c\mu + d) = b - \frac{ad}{c} = \frac{\det A}{-c}.$$

The strategy of part (a) can now be used, mutatis mutandis, to yield the desired result. $\square$

Theorem 5.7 (Smith form of Möbius transform).
Let $P(\lambda)$ be an $m \times n$ matrix polynomial over a field $\mathbb{F}$ with $k = \mathrm{grade}(P)$ and $r = \mathrm{rank}(P)$. Suppose

$$D(\lambda) = \left[\begin{array}{c|c} \begin{matrix} \lambda^{\alpha_1}p_1(\lambda) & & \\ & \ddots & \\ & & \lambda^{\alpha_r}p_r(\lambda) \end{matrix} & \\ \hline & 0 \end{array}\right]_{m \times n} \quad \text{with} \quad p_j(0) \neq 0 \ \text{for} \ 1 \leq j \leq r$$

is the Smith form of $P(\lambda)$, and $\mathcal{J}(P, \infty) = (\beta_1, \beta_2, \ldots, \beta_r)$. Then for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL(2, \mathbb{F})$ with associated Möbius function $\mathsf{m}_A(\mu) = \frac{a\mu + b}{c\mu + d}$ and Möbius transformation $\mathbf{M}_A$, the Smith form of the $m \times n$ matrix polynomial $\mathbf{M}_A(P)(\mu)$ is

$$\widetilde{D}_A(\mu) = \left[\begin{array}{c|c} \begin{matrix} \gamma_1(a\mu + b)^{\alpha_1}(c\mu + d)^{\beta_1}\mathbf{M}_A(p_1)(\mu) & & \\ & \ddots & \\ & & \gamma_r(a\mu + b)^{\alpha_r}(c\mu + d)^{\beta_r}\mathbf{M}_A(p_r)(\mu) \end{matrix} & \\ \hline & 0 \end{array}\right]_{m \times n}, \quad (5.5)$$

where the constants γ_j are chosen so that each invariant polynomial $\widetilde{d}_i(\mu)$ in $\widetilde{D}_A(\mu)$ is monic. Here each scalar Möbius transform $\mathbf{M}_A(p_j)$ is taken with respect to $\deg(p_j)$.

Proof. First we show that $\widetilde{D}_A(\mu)$ in (5.5) is a Smith form over the field $\mathbb{F}$ for *some* matrix polynomial. Since the diagonal entries $\widetilde{d}_j(\mu)$ of $\widetilde{D}_A(\mu)$ are monic polynomials in $\mathbb{F}[\mu]$ by definition, we only need to establish the divisibility chain property $\widetilde{d}_j(\mu) | \widetilde{d}_{j+1}(\mu)$ for $j = 1, \ldots, r-1$. Clearly we have $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_r$ and $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_r$, since $(\alpha_1, \alpha_2, \ldots, \alpha_r)$ and $(\beta_1, \beta_2, \ldots, \beta_r)$ are both partial multiplicity sequences for P. Also $\mathbf{M}_A(p_j) | \mathbf{M}_A(p_{j+1})$ follows from $p_j | p_{j+1}$ by Corollary 3.24. Thus $\widetilde{d}_j | \widetilde{d}_{j+1}$, and hence $\widetilde{D}_A$ *is* a Smith form.

It remains to show that $\widetilde{D}_A(\mu)$ is *the* Smith form of the *particular* matrix polynomial $\mathbf{M}_A(P)$. Recall from Remark 4.6 that a Smith form of a polynomial over a field $\mathbb{F}$ is completely determined by the finite part of its Jordan characteristic over the algebraic closure $\overline{\mathbb{F}}$. Thus the proof will be complete once we show that $\widetilde{D}_A$ and $\mathbf{M}_A(P)$ have the same finite Jordan characteristic, i.e.,

$$\mathcal{J}(\widetilde{D}_A, \mu_0) = \mathcal{J}(\mathbf{M}_A(P), \mu_0) \tag{5.6}$$

for all *finite* $\mu_0 \in \overline{\mathbb{F}}$.

Since the partial multiplicity sequences of any matrix polynomial are trivial for all but finitely many μ_0, (5.6) will automatically be satisfied for almost all $\mu_0 \in \overline{\mathbb{F}}$. Our strategy, therefore, is to focus on the sets

$$\mathcal{S}_{\widetilde{D}} := \left\{\mu_0 \in \overline{\mathbb{F}} : \mathcal{J}(\widetilde{D}_A, \mu_0) \text{ is nontrivial}\right\}$$
$$\text{and} \quad \mathcal{S}_{\mathbf{M}} := \left\{\mu_0 \in \overline{\mathbb{F}} : \mathcal{J}(\mathbf{M}_A(P), \mu_0) \text{ is nontrivial}\right\},$$

and show that (5.6) holds whenever $\mu_0 \in \mathcal{S}_{\widetilde{D}} \cup \mathcal{S}_{\mathbf{M}}$. Consequently, $\mathcal{S}_{\widetilde{D}} = \mathcal{S}_{\mathbf{M}}$ and (5.6) holds for *all* $\mu_0 \in \overline{\mathbb{F}}$.

We begin with $\mu_0 \in \mathcal{S}_{\widetilde{D}}$. In order to calculate partial multiplicity sequences of $\widetilde{D}_A$, first observe that in each diagonal entry $\widetilde{d}_j(\mu)$ of (5.5), the terms $(a\mu + b)^{\alpha_j}$, $(c\mu + d)^{\beta_j}$, and $\mathbf{M}_A(p_j)$ are pairwise relatively prime, and hence share no common roots. Consequently, contributions to any particular *nontrivial* partial multiplicity sequence for $\widetilde{D}_A$ can come either

- from just the $(a\mu + b)^{\alpha_j}$-terms,

- from just the $(c\mu + d)^{\beta_j}$-terms,

- or from just the $\mathbf{M}_A(p_j)$-terms.

That $(a\mu + b)$ is relatively prime to $(c\mu + d)$ follows from the nonsingularity of A. If either $(a\mu + b)$ or $(c\mu + d)$ were a factor of $\mathbf{M}_A(p_j)$, then by Lemma 5.6, $p_j(\lambda)$ would have to have either $\mathsf{m}_A(\frac{-b}{a})$ or $\mathsf{m}_A(\frac{-d}{c})$ as a root; however, neither of these is possible, since $\mathsf{m}_A(\frac{-b}{a}) = 0$ and $\mathsf{m}_A(\frac{-d}{c}) = \infty$.

Now consider the possible values of $\mu_0 \in \mathcal{S}_{\widetilde{D}}$. Clearly $\mu_0 = \frac{-b}{a}$ is one such possibility, but only if $a \neq 0$ and if $(\alpha_1, \alpha_2, \ldots, \alpha_r)$ is nontrivial; otherwise all the terms $(a\mu+b)^{\alpha_j}$ are nonzero constants, and are effectively absorbed by the normalizing constants γ_j. So if $\mu_0 = \frac{-b}{a} \in \mathcal{S}_{\widetilde{D}}$, then we have

$$\mathcal{J}\big(\widetilde{D}_A, \tfrac{-b}{a}\big) = (\alpha_1, \ldots, \alpha_r) = \mathcal{J}(D, 0) = \mathcal{J}(P, 0) = \mathcal{J}\big(\mathbf{M}_A(P), \mathsf{m}_A^{-1}(0)\big) = \mathcal{J}\big(\mathbf{M}_A(P), \tfrac{-b}{a}\big),$$

and (5.6) holds for $\mu_0 = \frac{-b}{a}$. Note that the fourth equality is an instance of Theorem 5.3, which we will continue to use freely throughout the remainder of this argument. Another possible μ_0-value in $\mathcal{S}_{\widetilde{D}}$ is $\mu_0 = \frac{-d}{c}$, but again only if $c \neq 0$ and if $(\beta_1, \beta_2, \ldots, \beta_r)$ is nontrivial; otherwise all $(c\mu + d)^{\beta_j}$-terms are nonzero constants that get absorbed by the constants γ_j. If $\mu_0 = \frac{-d}{c} \in \mathcal{S}_{\widetilde{D}}$, then we have

$$\mathcal{J}\big(\widetilde{D}_A, \tfrac{-d}{c}\big) = (\beta_1, \ldots, \beta_r) = \mathcal{J}(P, \infty) = \mathcal{J}\big(\mathbf{M}_A(P), \mathsf{m}_A^{-1}(\infty)\big) = \mathcal{J}\big(\mathbf{M}_A(P), \tfrac{-d}{c}\big),$$

and (5.6) holds for $\mu_0 = \frac{-d}{c}$. The only other μ_0-values in $\mathcal{S}_{\widetilde{D}}$ are the roots of the polynomial $\mathbf{M}_A(p_r)$, so suppose μ_0 is one of those roots. Then by Lemma 5.6 we have $\mathcal{J}\big(\widetilde{D}_A, \mu_0\big) = \mathcal{J}\big(D, \mathsf{m}_A(\mu_0)\big)$ with a finite $\mathsf{m}_A(\mu_0)$, and thus

$$\mathcal{J}\big(\widetilde{D}_A, \mu_0\big) = \mathcal{J}\big(D, \mathsf{m}_A(\mu_0)\big) = \mathcal{J}\big(P, \mathsf{m}_A(\mu_0)\big) = \mathcal{J}\big(\mathbf{M}_A(P), \mu_0\big),$$

completing the proof that (5.6) holds for all $\mu_0 \in \mathcal{S}_{\widetilde{D}}$.

Finally we turn to the μ_0-values in $\mathcal{S}_{\mathbf{M}}$. By Theorem 5.3 we certainly have

$$\mathcal{J}\big(\mathbf{M}_A(P), \mu_0\big) = \mathcal{J}\big(P, \mathsf{m}_A(\mu_0)\big)$$

for any $\mu_0 \in \mathcal{S}_{\mathbf{M}}$. The remainder of the argument depends on whether $\mathsf{m}_A(\mu_0)$ is finite or infinite.

(1) If $\mathsf{m}_A(\mu_0)$ is finite, then $\mathcal{J}\big(\mathbf{M}_A(P), \mu_0\big) = \mathcal{J}\big(P, \mathsf{m}_A(\mu_0)\big) = \mathcal{J}\big(D, \mathsf{m}_A(\mu_0)\big)$, since D is the Smith form of P. This partial multiplicity sequence is nontrivial since $\mu_0 \in \mathcal{S}_{\mathbf{M}}$, so either $\mathsf{m}_A(\mu_0) = 0$, or $\lambda_0 = \mathsf{m}_A(\mu_0) \neq 0$ is a root of $p_r(\lambda)$. If $\mathsf{m}_A(\mu_0) = 0$, then since μ_0 is finite we must have $a \neq 0$ and $\mu_0 = \frac{-b}{a}$, and hence $\mathcal{J}\big(D, \mathsf{m}_A(\mu_0)\big) = \mathcal{J}(D, 0) = (\alpha_1, \alpha_2, \ldots, \alpha_r) = \mathcal{J}\big(\widetilde{D}_A, \tfrac{-b}{a}\big) = \mathcal{J}\big(\widetilde{D}_A, \mu_0\big)$, so (5.6) holds. If on the other hand $\lambda_0 = \mathsf{m}_A(\mu_0) \neq 0$, then since $\lambda_0 = \mathsf{m}_A(\mu_0)$ is a root of $p_r(\lambda)$, we have by Lemma 5.6 that $\mathcal{J}\big(D, \mathsf{m}_A(\mu_0)\big) = \mathcal{J}\big(\widetilde{D}_A, \mu_0\big)$, and so once again (5.6) holds.

(2) If $\mathsf{m}_A(\mu_0) = \infty$, then because μ_0 is finite we must have $c \neq 0$ and $\mu_0 = \frac{-d}{c}$. In this case $\mathcal{J}(\mathbf{M}_A(P), \mu_0) = \mathcal{J}(P, \infty) = (\beta_1, \beta_2, \ldots, \beta_r)$ by our choice of notation; this sequence is nontrivial since $\mu_0 \in \mathcal{S}_{\mathbf{M}}$. But $(\beta_1, \beta_2, \ldots, \beta_r)$ is equal to $\mathcal{J}(\widetilde{D}_A, \frac{-d}{c})$ when $c \neq 0$, by our definition of $\widetilde{D}_A$, and so (5.6) holds for this final case.

Thus (5.6) holds for *all* $\mu_0 \in \mathcal{S}_{\mathbf{M}}$, and the proof is complete. □

Using the fact that the reversal operator can be viewed as a Möbius transformation, an immediate corollary of Theorem 5.7 is the complete characterization of the relationship between the Smith forms of P and $\mathrm{rev}_k P$, as described earlier (without proof) in Theorem 4.17.

6 Möbius and Invariant Pairs

It is common knowledge, see e.g., [13, 50] for the case of Hamiltonian and symplectic matrices and pencils, that invariant subspaces of matrices and deflating subspaces of matrix pencils remain invariant under Cayley transformations. In this section, we investigate the analogous question for regular matrix polynomials under general Möbius transformations, using the concept of *invariant pair*. Introduced in [10] and further developed in [9], this notion extends the well-known concepts of *standard pair* [30] and *null pair* [8]. Just as invariant subspaces for matrices or deflating subspaces for pencils can be seen as a generalization of eigenvectors, we can think of invariant pairs of matrix polynomials as a generalization of *eigenpairs* (x, λ_0), consisting of an eigenvector x together with its associated eigenvalue λ_0. The concept of invariant pair can be a more flexible and useful tool for computation in the context of matrix polynomials than either eigenpair, null pair, or standard pair [9].

Definition 6.1 (Invariant pair).
Let $P(\lambda) = \sum_{j=0}^{k} \lambda^j B_j$ be a regular $n \times n$ matrix polynomial of grade k over a field $\mathbb{F}$, and let $(X, S) \in \mathbb{F}^{n \times m} \times \mathbb{F}^{m \times m}$. Then (X, S) is said to be an invariant pair *for $P(\lambda)$ if the following two conditions hold:*

(a) $P(X, S) := \sum_{j=0}^{k} B_j X S^j = 0$, *and*

(b) $\mathcal{V}_k(X, S) := \begin{bmatrix} X \\ XS \\ \vdots \\ XS^{k-1} \end{bmatrix}$ *has full column rank.*

Example 6.2. The following provides a variety of examples of invariant pairs:

(a) The simplest example of an invariant pair is an eigenpair: with $m = 1$, a vector-scalar pair (x, λ_0) with nonzero $x \in \mathbb{F}^n$ and $\lambda_0 \in \mathbb{F}$ is an invariant pair for P if and only if (x, λ_0) is an eigenpair for P.

(b) Standard pairs in the sense of [30] are the same as invariant pairs of monic matrix polynomials over $\mathbb{F} = \mathbb{C}$ with $k = \deg P$ and $m = nk$.

(c) If $\mathbb{F} = \mathbb{C}$ and S is a single Jordan block associated with the eigenvalue λ_0, then the columns of the matrix X in any invariant pair (X, S) form a *Jordan chain* for $P(\lambda)$ associated with the eigenvalue λ_0. It is worth noting that Jordan chains constitute an alternative to the Smith form as a means of defining the Jordan characteristic of a matrix polynomial at an eigenvalue $\lambda_0 \in \mathbb{F}$. See [30] for more on Jordan chains.

(d) If $\mathbb{F} = \mathbb{C}$, then a right null pair for P (in the sense of [8]) associated with $\lambda_0 \in \mathbb{C}$ is the same as an invariant pair (X, S), where the spectrum of S is just $\{\lambda_0\}$ and the size m coincides with the algebraic multiplicity of λ_0 as an eigenvalue of P.

(e) For regular *pencils*, invariant pairs are closely related to the notion of *deflating subspace*.

Recall that for an $n \times n$ pencil $L(\lambda) = \lambda A + B$, an m-dimensional subspace $\mathcal{X} \subseteq \mathbb{F}^n$ is a deflating subspace for $L(\lambda)$, see e.g., [11, 39, 60], if there exists another m-dimensional subspace $\mathcal{Y} \in \mathbb{F}^n$ such that $A\mathcal{X} \subseteq \mathcal{Y}$ and $B\mathcal{X} \subseteq \mathcal{Y}$. Letting the columns of $X, Y \in \mathbb{F}^{n \times m}$ form bases for $\mathcal{X}$ and $\mathcal{Y}$, respectively, it follows that these containment relations can be equivalently expressed as $AX = YW_A$ and $BX = YW_B$, for some matrices $W_A, W_B \in \mathbb{F}^{m \times m}$.

Now if $A\mathcal{X} = \mathcal{Y}$, i.e., if $\mathcal{X} \cap \ker A = \{0\}$, then W_A is invertible, so that $Y = AXW_A^{-1}$ and $BX = AXW_A^{-1}W_B$. Letting $S := -W_A^{-1}W_B$, we see that $L(X, S) = AXS + BX = 0$, and hence (X, S) is an invariant pair for $L(\lambda)$. Similarly if $\mathcal{X} \cap \ker B = \{0\}$, then W_B is invertible, and $(X, \widetilde{S})$ with $\widetilde{S} := -W_B^{-1}W_A$ is an invariant pair for the pencil $\mathrm{rev}\,L(\lambda) = \lambda B + A$. Later (in Definition 6.7) we will refer to such a pair $(X, \widetilde{S})$ as a reverse invariant pair for $L(\lambda)$.

Conversely, if (X, S) is an invariant pair for $L(\lambda)$, then $\mathcal{X}$, the subspace spanned by the columns of X, will be a deflating subspace for $L(\lambda)$. To see why this is so, first observe that $L(X, S) = AXS + BX = 0$ implies that $B\mathcal{X} \subseteq A\mathcal{X}$. Thus if $\dim \mathcal{X} = m$, then *any* m-dimensional subspace $\mathcal{Y} \in \mathbb{F}^n$ such that $A\mathcal{X} \subseteq \mathcal{Y}$ will witness that $\mathcal{X}$ is a deflating subspace for L.

Remark 6.3. Definition 6.1 follows the terminology in [10], where both conditions (a) and (b) are required for a pair (X, S) to be an invariant pair for P. A weaker notion is introduced in [9], where condition (a) alone suffices for (X, S) to be called an invariant pair; if conditions (a) and (b) are both met, then (X, S) is called a minimal invariant pair. This weaker notion of invariant pair has several drawbacks that significantly detract from its usefulness. First, weak invariant pairs may contain redundant information, and even be arbitrarily large; for example, if (x, λ_0) is any eigenpair for P, then $X = [\,x \ x \ \ldots \ x\,]_{n \times m}$ with $S = \lambda_0 I_m$ gives weak invariant pairs for P of unbounded size. The full column rank condition (b), though, forces $m \leq kn$ since $\mathcal{V}_k(X, S)$ has only kn rows. A more significant drawback of weak invariant pairs (X, S) concerns the spectrum $\sigma(S)$ and its relationship to $\sigma(P)$. The intended purpose of S is to capture some portion of the eigenvalues of $P(\lambda)$, i.e., $\sigma(S) \subseteq \sigma(P)$, while eigenvector information corresponding to this subset $\sigma(S)$ is captured in X. We will see in Proposition 6.6 that any invariant pair in the sense of Definition 6.1 does indeed have the desired containment property $\sigma(S) \subseteq \sigma(P)$. However, a weak invariant pair may include many *spurious eigenvalues*, that is, eigenvalues of S that are *not* in $\sigma(P)$. In particular, a weak invariant pair (X, S) with nonzero X and $m \times m$ matrix S may have up to $m - 1$ spurious eigenvalues; in general S and $P(\lambda)$ can only be guaranteed to have one eigenvalue in common. For these reasons we have reserved the more concise term invariant pair for the more important notion, as specified in Definition 6.1.

For polynomials of the special form $P(\lambda) = \lambda I_n - B_0$, the condition in Definition 6.1(a) reduces to $P(X, S) = XS - B_0X = 0$, equivalently to $B_0X = XS$, which is satisfied if and only if the columns of X span an invariant subspace $\mathcal{X}$ of the matrix B_0. In this case, condition (b) just means that the columns of X are linearly independent, and thus form a basis for $\mathcal{X}$. The eigenvalues of S are then the eigenvalues of B_0 associated with the subspace $\mathcal{X}$. Certainly an invariant subspace is independent of the choice of basis, and so can also be represented by the columns of any matrix XT, where T is an invertible $m \times m$ transformation matrix. This observation motivates the following definition.

Definition 6.4 (Similarity of pairs).
Two pairs $(X, S), (\widetilde{X}, \widetilde{S}) \in \mathbb{F}^{n \times m} \times \mathbb{F}^{m \times m}$, where $\mathbb{F}$ is an arbitrary field, are said to be similar *if there exists a nonsingular matrix $T \in \mathbb{F}^{m \times m}$ such that $\widetilde{X} = XT$ and $\widetilde{S} = T^{-1}ST$.*

The following basic property of similar pairs is straightforward to prove from the definitions.

Lemma 6.5. *If the pairs (X, S) and $(\widetilde{X}, \widetilde{S})$ are similar, then one is an invariant pair for a regular matrix polynomial $P(\lambda)$ if and only if the other one is, too.*

This simple observation has an important consequence: if (X, S) is an invariant pair for $P(\lambda)$, then every eigenvalue of the matrix S is also an eigenvalue of the matrix polynomial $P(\lambda)$, i.e., S has no spurious eigenvalues as discussed in Remark 6.3. More generally, any invariant subspace for S induces an invariant pair for $P(\lambda)$ that is in a certain sense "contained" in (X, S); this is made precise in the following proposition. Here for a matrix Y, we use $\mathrm{Col}(Y)$ to denote the space spanned by the columns of Y.

Proposition 6.6. *Let $(X, S) \in \mathbb{F}^{n \times m} \times \mathbb{F}^{m \times m}$ be an invariant pair for the regular matrix polynomial $P(\lambda) = \sum_{j=0}^{k} \lambda^j B_j$.*

(a) *Suppose that $\mathcal{U} \subseteq \mathbb{F}^m$ is an ℓ-dimensional invariant subspace for the matrix S, and $T = \begin{bmatrix} T_1 & T_2 \end{bmatrix} \in \mathbb{F}^{m \times m}$ is any nonsingular matrix such that the columns of T_1 form a basis for $\mathcal{U}$, so that*

$$\widetilde{S} := T^{-1}ST = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix}, \quad \text{where } S_{11} \in \mathbb{F}^{\ell \times \ell}.$$

Then (XT_1, S_{11}) is an invariant pair for $P(\lambda)$ that is "contained" in (X, S) in the sense that
$$\mathrm{Col}(XT_1) \subseteq \mathrm{Col}(X) \quad \text{and} \quad \sigma(S_{11}) \subseteq \sigma(S).$$

(b) *$\sigma(S) \subseteq \sigma(P)$.*

Proof. (a) First observe that $(XT, \widetilde{S})$ is an invariant pair for $P(\lambda)$, since it is similar to the invariant pair (X, S). But $\widetilde{S}^j = \begin{bmatrix} S_{11}^j & * \\ 0 & * \end{bmatrix}$ for all $j \geq 0$, so we have

$$XT\widetilde{S}^j = X \begin{bmatrix} T_1 & T_2 \end{bmatrix} \begin{bmatrix} S_{11}^j & * \\ 0 & * \end{bmatrix} = \begin{bmatrix} XT_1 S_{11}^j & * \end{bmatrix} \tag{6.1}$$

for all $j \geq 0$. Hence

$$0 = P(XT, \widetilde{S}) = \sum_{j=0}^{k} B_j XT\widetilde{S}^j = \sum_{j=0}^{k} \begin{bmatrix} B_j(XT_1)S_{11}^j & * \end{bmatrix} = \begin{bmatrix} P(XT_1, S_{11}) & * \end{bmatrix},$$

and so $P(XT_1, S_{11}) = 0$, i.e., the pair (XT_1, S_{11}) satisfies condition (a) of Definition 6.1. The containments $\mathrm{Col}(XT_1) \subseteq \mathrm{Col}(X)$ and $\sigma(S_{11}) \subseteq \sigma(S)$ are immediate; all that remains is to see why the full column rank condition holds for (XT_1, S_{11}). From (6.1) we see that $\mathcal{V}_k(XT, \widetilde{S}) = \begin{bmatrix} \mathcal{V}_k(XT_1, S_{11}) & * \end{bmatrix}$, so the full column rank of $\mathcal{V}_k(XT_1, S_{11})$ follows from the full column rank of $\mathcal{V}_k(XT, \widetilde{S})$.

(b) Suppose that $\lambda_0 \in \sigma(S)$, with corresponding eigenvector $y \in \mathbb{F}^m$. Then by part (a) with $\ell = 1$ we know that (Xy, λ_0) is an invariant pair for P. Thus (Xy, λ_0) is an eigenpair for P by Example 6.2, and so $\lambda_0 \in \sigma(P)$. $\qquad\square$

This proposition shows that invariant pairs enable us to focus on any specific part of the finite spectrum of $P(\lambda)$. Unfortunately, though, invariant pairs can never contain any information about eigenvalues at ∞. To address this deficiency, we introduce the following concept, which builds on the correspondence between the eigenvalue ∞ for $P(\lambda)$ and the eigenvalue zero for $\operatorname{rev} P(\lambda)$.

Definition 6.7 (Reverse invariant pairs).
Let $P(\lambda) = \sum_{j=0}^{k} \lambda^j B_j$ be a regular $n \times n$ matrix polynomial with grade k over a field $\mathbb{F}$. Then a pair $(X, S) \in \mathbb{F}^{n \times m} \times \mathbb{F}^{m \times m}$ is called a reverse invariant pair *for $P(\lambda)$ if (X, S) is an invariant pair for $\operatorname{rev}_k P(\lambda)$, i.e., if*

$$\operatorname{rev}_k P(X, S) := \sum_{j=0}^{k} B_{k-j} X S^j = 0 \,,$$

and $\mathcal{V}_k(X, S)$ as in Definition 6.1(b) has full column rank.

The next example and proposition provide two simple sources of reverse invariant pairs.

Example 6.8. A nonzero vector $x \in \mathbb{F}^n$ is an eigenvector associated with the eigenvalue ∞ of $P(\lambda)$ if and only if $(x, 0)$ is a reverse invariant pair for $P(\lambda)$.

Recall from Theorem 4.15 that reversal has a simple effect on nonzero eigenvalues and their Jordan characteristics, i.e., $\mathcal{J}(P, \lambda_0) \equiv \mathcal{J}(\operatorname{rev}_k P, 1/\lambda_0)$ for any nonzero λ_0. This suggests the following connection between invariant pairs and reverse invariant pairs for P.

Proposition 6.9. *Suppose $P(\lambda) = \sum_{j=0}^{k} \lambda^j B_j$ is a regular $n \times n$ matrix polynomial with grade k over a field $\mathbb{F}$, and let $(X, S) \in \mathbb{F}^{n \times m} \times \mathbb{F}^{m \times m}$ be a matrix pair with* nonsingular *S. Then (X, S) is an invariant pair for P if and only if (X, S^{-1}) is a reverse invariant pair for P.*

Proof. First observe that

$$P(X, S) \cdot S^{-k} = \sum_{j=0}^{k} B_j X S^{j-k} = \sum_{j=0}^{k} B_{k-j} X S^{-j} = \operatorname{rev}_k P(X, S^{-1})$$

and

$$\mathcal{V}_k(X, S) \cdot S^{-(k-1)} = (R_k \otimes I_n) \cdot \mathcal{V}_k(X, S^{-1}), \quad \text{where} \quad R_k := \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix}_{k \times k}$$

is the $k \times k$ reverse identity. From these relations it is clear that $P(X, S) = 0$ if and only if $\operatorname{rev}_k P(X, S^{-1}) = 0$, and that $\mathcal{V}_k(X, S)$ has full column rank if and only if $\mathcal{V}_k(X, S^{-1})$ does, too. The desired conclusion now follows. $\square$

We are now in a position to address the main issue of this section — how Möbius transformations interact with invariant pairs. In particular, given an invariant pair (X, S) for P, is there a naturally corresponding invariant pair for the Möbius transform $\mathbf{M}_A(P)$? From Theorem 5.3 we know that an eigenvalue μ_0 of P corresponds to an eigenvalue $\mathsf{m}_{A^{-1}}(\mu_0)$ of $\mathbf{M}_A(P)$. But Proposition 6.6 shows that $\sigma(S) \subseteq \sigma(P)$ for any invariant pair (X, S), and hence $\mathsf{m}_{A^{-1}}(\sigma(S)) \subseteq \sigma(\mathbf{M}_A(P))$. Now as long as the matrix $\mathsf{m}_{A^{-1}}(S)$ is *defined*, then $\sigma(\mathsf{m}_{A^{-1}}(S)) = \mathsf{m}_{A^{-1}}(\sigma(S))$, and hence $\sigma(\mathsf{m}_{A^{-1}}(S)) \subseteq \sigma(\mathbf{M}_A(P))$. Thus it is reasonable to conjecture that an invariant pair (X, S) for P will induce an invariant pair $(X, \mathsf{m}_{A^{-1}}(S))$ for the Möbius transform $\mathbf{M}_A(P)$, whenever the matrix $\mathsf{m}_{A^{-1}}(S)$ is defined. This conjecture, along with several related variations, is proved in Theorem 6.11. We begin with the following lemma, whose straightforward proof is only sketched.

Lemma 6.10. *Suppose $\phi_j(\lambda) \in \mathbb{F}[\lambda]$ for $j = 0, \ldots, k$ are arbitrary scalar polynomials over $\mathbb{F}$, and $P(\lambda) = \sum_{j=0}^{k} \phi_j(\lambda) B_j$ is any $n \times n$ matrix polynomial expressed in terms of the ϕ_j's. If $(X, S) \in F^{n \times m} \times \mathbb{F}^{m \times m}$ is any matrix pair, then $P(X, S) = \sum_{j=0}^{k} B_j X \phi_j(S)$.*

Proof. First show that the additive property $(P+Q)(X, S) = P(X, S) + Q(X, S)$ holds for any polynomials P and Q of the same size, by expressing P and Q in the standard basis and then using the defining formula given in Definition 6.1(a). Then show that $P(X, S) = BX\phi(S)$ for any polynomial of the special form $P(\lambda) = \phi(\lambda)B$. Combining the additive property with the special case gives the desired result. $\qquad\square$

Theorem 6.11 (Invariant pairs of Möbius transforms)**.**
Let $P(\lambda) = \sum_{j=0}^{k} \lambda^j B_j$ be an $n \times n$ matrix polynomial over a field $\mathbb{F}$ and let (X, S) be a matrix pair in $\mathbb{F}^{n \times m} \times \mathbb{F}^{m \times m}$. Furthermore, let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL(2, \mathbb{F}), \quad \mathsf{m}_A(\lambda) = \frac{a\lambda + b}{c\lambda + d}, \quad and \quad R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

(a) *Suppose that either $c = 0$, or $c \neq 0$ and $\frac{a}{c} \notin \sigma(S)$. Then*
    (X, S) *is an invariant pair for* P $\iff$ $(X, \mathsf{m}_{A^{-1}}(S))$ *is an invariant pair for* $\mathbf{M}_A(P)$.

(b) *Suppose that either $d = 0$, or $d \neq 0$ and $\frac{b}{d} \notin \sigma(S)$. Then*
    (X, S) *is an invariant pair for* P $\iff$ $(X, \mathsf{m}_{(AR)^{-1}}(S))$ *is a reverse invariant pair for* $\mathbf{M}_A(P)$.

(c) *Suppose that either $a = 0$, or $a \neq 0$ and $\frac{c}{a} \notin \sigma(S)$. Then*
    (X, S) *is a reverse invariant pair for* P $\iff$ $(X, \mathsf{m}_{(RA)^{-1}}(S))$ *is an invariant pair for* $\mathbf{M}_A(P)$.

(d) *Suppose that either $b = 0$, or $b \neq 0$ and $\frac{d}{b} \notin \sigma(S)$. Then*
    (X, S) *is a reverse invariant pair for* P $\iff$ $(X, \mathsf{m}_{(RAR)^{-1}}(S))$ *is a reverse invariant pair for* $\mathbf{M}_A(P)$.

Proof. (a) The given hypothesis on A and S (i.e., either that $c = 0$ (and hence $a, d \neq 0$), or that $c \neq 0$ with $\frac{a}{c}$ not an eigenvalue of S) is exactly the condition needed to guarantee that the matrix $(-cS + aI)$ is nonsingular, and thus that $\widetilde{S} := \mathsf{m}_{A^{-1}}(S) = (dS - bI)(-cS + aI)^{-1}$ introduced in (3.8) is well defined. Now $\mathsf{m}_A(\frac{-d}{c}) = \infty$, so the invertibility of Möbius functions implies that $\mathsf{m}_{A^{-1}}(\infty) = \frac{-d}{c}$. But ∞ is not an eigenvalue of the matrix S, so $\frac{-d}{c}$ is not an eigenvalue of $\mathsf{m}_{A^{-1}}(S) = \widetilde{S}$; thus under these conditions the matrix $(c\widetilde{S} + dI)$ is also nonsingular, no matter whether c is zero or nonzero. Consequently, $\mathsf{m}_A(\widetilde{S}) = (a\widetilde{S} + bI)(c\widetilde{S} + dI)^{-1}$ is well defined and equal to S.

To show the equivalence of (X, S) and $(X, \widetilde{S})$ being invariant pairs, we consider the two defining conditions in turn. A direct calculation using Lemma 6.10 first establishes a general relationship between $\mathbf{M}_A(P)(X, \widetilde{S})$ and $P(X, S)$:

$$
\begin{aligned}
\mathbf{M}_A(P)(X, \widetilde{S}) &= \sum_{j=0}^{k} B_j X (a\widetilde{S} + bI)^j (c\widetilde{S} + dI)^{k-j} \\
&= \left(\sum_{j=0}^{k} B_j X \big((a\widetilde{S} + bI)(c\widetilde{S} + dI)^{-1} \big)^j \right) \cdot (c\widetilde{S} + dI)^k \\
&= P\big(X, \mathsf{m}_A(\widetilde{S})\big) \cdot (c\widetilde{S} + dI)^k \\
&= P(X, S) \cdot (c\widetilde{S} + dI)^k.
\end{aligned}
$$

Then the nonsingularity of $(c\widetilde{S}+dI)$ implies that $\mathbf{M}_A(P)(X,\widetilde{S}) = 0$ if and only if $P(X,S) = 0$, so Definition 6.1(a) is satisfied either for both pairs or for neither of them.

For condition (b) in Definition 6.1, we first see how $\mathcal{V}_k(X,S)$ and $\mathcal{V}_k(X,\widetilde{S})$ are related. Since $S = \mathsf{m}_A(\widetilde{S})$, we have

$$
\mathcal{V}_k(X,S)\cdot(c\widetilde{S}+dI)^{k-1} = \begin{bmatrix} X \\ X\mathsf{m}_A(\widetilde{S}) \\ \vdots \\ X\mathsf{m}_A(\widetilde{S})^{k-1} \end{bmatrix}(c\widetilde{S}+dI)^{k-1} = \begin{bmatrix} Xq_1(\widetilde{S}) \\ Xq_2(\widetilde{S}) \\ \vdots \\ Xq_k(\widetilde{S}) \end{bmatrix}, \qquad (6.2)
$$

where each q_i is a scalar polynomial of degree at most $k-1$. Assembling all the coefficients of the q_i polynomials (row-wise) into a $k \times k$ matrix $\widetilde{\mathcal{Q}}$, we see that the last matrix in (6.2) can be expressed as $(\widetilde{\mathcal{Q}} \otimes I_n)\mathcal{V}_k(X,\widetilde{S})$, and so

$$
(\widetilde{\mathcal{Q}} \otimes I_n)\cdot\mathcal{V}_k(X,\widetilde{S}) = \mathcal{V}_k(X,S)\cdot(c\widetilde{S}+dI)^{k-1} \qquad (6.3)
$$

gives a general relationship between $\mathcal{V}_k(X,\widetilde{S})$ and $\mathcal{V}_k(X,S)$. The nonsingularity of $(c\widetilde{S}+dI)$ now implies that

$$
\operatorname{rank}\mathcal{V}_k(X,\widetilde{S}) \ge \operatorname{rank}\mathcal{V}_k(X,S).
$$

Interchanging the roles of S and $\widetilde{S}$, and using that $\widetilde{S} = \mathsf{m}_{A^{-1}}(S)$, a similar argument shows that

$$
(\mathcal{Q} \otimes I_n)\cdot\mathcal{V}_k(X,S) = \mathcal{V}_k(X,\widetilde{S})\cdot(-cS+aI)^{k-1}
$$

for some $k \times k$ matrix of coefficients $\mathcal{Q}$. Then the nonsingularity of $(-cS+aI)$ implies that

$$
\operatorname{rank}\mathcal{V}_k(X,\widetilde{S}) \le \operatorname{rank}\mathcal{V}_k(X,S),
$$

and thus $\operatorname{rank}\mathcal{V}_k(X,\widetilde{S}) = \operatorname{rank}\mathcal{V}_k(X,S)$. Consequently Definition 6.1(b) is satisfied either for both or for neither of the two pairs. This completes the proof of part (a).

(b) Consider the Möbius transformation induced by the matrix $B = AR = \left[\begin{smallmatrix} b & a \\ d & c \end{smallmatrix}\right]$. Then the given hypothesis on d and b allows us to use part (a) to conclude that

$$(X,S) \text{ is an invariant pair for } P \iff (X,\mathsf{m}_{B^{-1}}(S)) \text{ is an invariant pair for } \mathbf{M}_B(P).$$

But an invariant pair for $\mathbf{M}_B(P) = \mathbf{M}_{AR}(P) = \mathbf{M}_R\big(\mathbf{M}_A(P)\big) = \operatorname{rev}_k \mathbf{M}_A(P)$ is a reverse invariant pair for $\mathbf{M}_A(P)$, so the result is proved.

(c) By definition, (X,S) is a reverse invariant pair for P if and only if (X,S) is an invariant pair for $\operatorname{rev}_k P$. Applying (a) to the polynomial $\operatorname{rev}_k P$, with Möbius transformation induced by the matrix $RA = \left[\begin{smallmatrix} c & d \\ a & b \end{smallmatrix}\right]$, we obtain that (X,S) is an invariant pair for $\operatorname{rev}_k P$ if and only if $(X,\mathsf{m}_{(RA)^{-1}}(S))$ is an invariant pair for $\mathbf{M}_{RA}(\operatorname{rev}_k P) = \mathbf{M}_A\big(\mathbf{M}_R(\operatorname{rev}_k P)\big) = \mathbf{M}_A(P)$, and the result is proved.

(d) The argument is analogous to that for (c). Apply (a) to the polynomial $\operatorname{rev}_k P$, using the Möbius transformation induced by the matrix $RAR = \left[\begin{smallmatrix} d & c \\ b & a \end{smallmatrix}\right]$. Then we obtain that (X,S) is an invariant pair for $\operatorname{rev}_k P$ if and only if $(X,\mathsf{m}_{(RAR)^{-1}}(S))$ is an invariant pair for $\mathbf{M}_{RAR}(\operatorname{rev}_k P) = \mathbf{M}_R\big(\mathbf{M}_A\big(\mathbf{M}_R(\operatorname{rev}_k P)\big)\big) = \operatorname{rev}_k \mathbf{M}_A(P)$. $\qquad\square$

Remark 6.12. From Theorem 6.11, it is straightforward to show that if (x,λ_0) is any eigenpair of the regular matrix polynomial P, including any eigenpair with $\lambda_0 = \infty$, then $(x,\mathsf{m}_{A^{-1}}(\lambda_0))$ will always be an eigenpair of $\mathbf{M}_A(P)$. Thus we see that *eigenvectors of matrix polynomials are preserved by any Möbius transformation.* Are Jordan chains also preserved by Möbius transformations? Approaching this question via the usual definition of Jordan chains [30] leads to technical difficulties; the invariant pair point of view, though, provides a

clearer view of the situation. Recall from Example 6.2(c) that if (X, S) is an invariant pair for $P(\lambda)$ and $S = \mathcal{J}_m(\lambda_0)$ is the $m \times m$ Jordan block associated with the eigenvalue λ_0, then the columns of X form a Jordan chain for $P(\lambda)$ associated with the eigenvalue λ_0. Now by Theorem 6.11 we know that $(X, \mathsf{m}_{A^{-1}}(S))$ is the corresponding invariant pair for $\mathbf{M}_A(P)$, as long as $\mathsf{m}_{A^{-1}}(S)$ is defined. But the matrix $\mathsf{m}_{A^{-1}}(S)$ will in general not be in Jordan form, although it will always be upper triangular and Toeplitz, with $\mathsf{m}_{A^{-1}}(\lambda_0)$ on the diagonal. Consequently the columns of X will usually *not* form a Jordan chain for $\mathbf{M}_A(P)$ associated with the eigenvalue $\mathsf{m}_{A^{-1}}(\lambda_0)$. One notable exception is when $m = 2$; in this case the 2×2 matrix $\mathsf{m}_{A^{-1}}(S)$ is always close enough to Jordan form that a simple scaling suffices to recover a Jordan chain of length two for $\mathbf{M}_A(P)$ from the two columns of X. In summary, then, we see that *Jordan chains are not in general preserved under Möbius transformations*. Thus it is reasonable to expect that invariant pairs will usually be a more natural and convenient tool than Jordan chains in situations where Möbius transformations are involved.

An intrinsic limitation of any invariant pair of a matrix polynomial P is that it can only contain information about the finite spectrum of P, since that spectral information is recorded in a *matrix S*. A reverse invariant pair for P overcomes this limitation, but at the cost of not being able to contain information about the eigenvalue 0. These limitations make it clear, for example, why the notion of a deflating subspace of a regular pencil, as described in Example 6.2(e), is (for pencils) more general than either invariant pair or reverse invariant pair; deflating subspaces have no trouble in simultaneously accommodating both zero and infinite eigenvalues.

The "forbidden eigenvalue" phenomenon of Theorem 6.11 is another setting in which these intrinsic limitations of (reverse) invariant pairs manifest themselves. For example, suppose we have an invariant pair (X, S) for a matrix polynomial P, and we wish to find an invariant pair with the same first component X for a Möbius transform $\mathbf{M}_A(P)$. Since such an invariant pair for $\mathbf{M}_A(P)$ would have to include all of $\mathsf{m}_{A^{-1}}(\sigma(S))$ among its associated eigenvalues, we must make sure that $\mathsf{m}_{A^{-1}}(\sigma(S))$ does *not* contain ∞, equivalently that $\mathsf{m}_A(\infty) \notin \sigma(S)$. This $\mathsf{m}_A(\infty)$ is the "forbidden eigenvalue" $\frac{a}{c}$ of Theorem 6.11(a). Analogous considerations account for the "forbidden eigenvalues" in each of the other parts of Theorem 6.11.

Suppose now that we have an invariant pair (X, S) for P, where S contains *both* of the forbidden eigenvalues in parts (a) and (b) of Theorem 6.11. Then it seems that nothing can be said about the effect of the Möbius transformation $\mathbf{M}_A$ on the invariant pair; Theorem 6.11 cannot be used to find either an invariant or a reverse invariant pair for $\mathbf{M}_A(P)$ with first component X. However, if we can "separate" the two forbidden eigenvalues from each other, then it is possible to say something useful about the effect of $\mathbf{M}_A$ on the invariant pair (X, S). This procedure is illustrated in the following remark. An analogous procedure can be formulated if we start instead with a reverse invariant pair for P.

Remark 6.13. Let $P(\lambda) = \sum_{j=0}^{k} \lambda^j B_j$ be a regular $n \times n$ matrix polynomial over a field $\mathbb{F}$, and let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL(2, \mathbb{F}) \quad \text{and} \quad \mathsf{m}_A(\lambda) = \frac{a\lambda + b}{c\lambda + d}, \quad \text{with} \quad a, c, d \neq 0.$$

Suppose $(X, S) \in F^{n \times m} \times \mathbb{F}^{m \times m}$ is an invariant pair for P with $\frac{a}{c}, \frac{b}{d} \in \sigma(S)$, so that the spectrum of S contains the two "forbidden" eigenvalues from Theorem 6.11(a) and (b). Observe that $\frac{a}{c} = \mathsf{m}_A(\infty)$ and $\frac{b}{d} = \mathsf{m}_A(0)$ must be distinct elements of $\mathbb{F}$, since m_A is a bijection on $\mathbb{F}_\infty$. Now there exists a nonsingular matrix $T \in \mathbb{F}^{m \times m}$ such that

$$\widetilde{X} := XT = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \quad \text{and} \quad \widetilde{S} := T^{-1}ST = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix},$$

where $X_1 \in \mathbb{F}^{n \times \ell}$, $X_2 \in \mathbb{F}^{n \times (m-\ell)}$, $S_1 \in \mathbb{F}^{\ell \times \ell}$, $S_2 \in \mathbb{F}^{(m-\ell) \times (m-\ell)}$ for some $1 \leq \ell < m$, and where $\frac{a}{c} \notin \sigma(S_1)$ and $0, \frac{b}{d} \notin \sigma(S_2)$. Thus the eigenvalues $\frac{b}{d}$ and $\frac{a}{c}$ are separated into

distinct blocks S_1 and S_2, respectively, where S_2 can even be chosen to be invertible. (One possible way to do this separation is to bring S into rational canonical form [37, Sect. 3.4] over the field $\mathbb{F}$, with one Frobenius companion block for each $\mathbb{F}$-elementary divisor of S; then gather all companion blocks associated with eigenvalues 0 and $\frac{b}{d}$ in S_1, all companion blocks associated with $\frac{a}{c}$ in S_2, and distribute all other companion blocks arbitrarily among S_1 and S_2.) Then $(\widetilde{X}, \widetilde{S})$ is an invariant pair for P by Lemma 6.5, and (X_1, S_1) and (X_2, S_2) are invariant pairs for P by Proposition 6.6. Furthermore, Theorem 6.11 can now be applied to (X_1, S_1) and (X_2, S_2) individually to conclude that $(X_1, \mathsf{m}_{A^{-1}}(S_1))$ is an invariant pair for $\mathbf{M}_A(P)$ and $(X_2, \mathsf{m}_{(AR)^{-1}}(S_2))$ is a reverse invariant pair for $\mathbf{M}_A(P)$.

An alternative approach to this situation is to organize the data X_1, X_2, S_1, S_2 from the previous paragraph in a slightly different way. We still view (X_1, S_1) as an invariant pair for P, but now note that (X_2, S_2^{-1}) is a reverse invariant pair for P, by the discussion in Example 6.8(b). Then consider the pair $(\widehat{X}, \widehat{S})$, where

$$\widehat{X} := \widetilde{X} = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \quad \text{and} \quad \widehat{S} := \begin{bmatrix} S_1 & 0 \\ 0 & S_2^{-1} \end{bmatrix};$$

clearly $(\widehat{X}, \widehat{S})$ will in general be neither an invariant pair nor a reverse invariant pair for P. However, the matrix

$$\mathcal{W}_k\big((X_1, S_1), (X_2, S_2^{-1})\big) := \begin{bmatrix} X_1 & X_2(S_2^{-1})^{k-1} \\ X_1 S_1 & X_2(S_2^{-1})^{k-2} \\ \vdots & \vdots \\ X_1 S_1^{k-1} & X_2 \end{bmatrix} = \mathcal{V}_k(\widetilde{X}, \widetilde{S}) \cdot \begin{bmatrix} I_\ell & 0 \\ 0 & (S_2^{-1})^{k-1} \end{bmatrix} \tag{6.4}$$

has full column rank, since $(\widetilde{X}, \widetilde{S})$ being an invariant pair implies that $\mathcal{V}_k(\widetilde{X}, \widetilde{S})$ has full column rank. Thus the pair $(\widehat{X}, \widehat{S})$ can be viewed as a natural generalization of the notion of a *decomposable pair* for P, as introduced in [30]. Such a "generalized decomposable pair" consists of an invariant pair together with a reverse invariant pair, satisfying a full rank condition for a matrix $\mathcal{W}_k$ as in (6.4).

Now since $\frac{a}{c} \notin \sigma(S_1)$ and $\frac{d}{b} \notin \sigma(S_2^{-1})$, we can apply Theorem 6.11 to the individual pairs (X_1, S_1) and (X_2, S_2^{-1}) to conclude that $(X_1, \mathsf{m}_{A^{-1}}(S_1))$ is an invariant pair for $\mathbf{M}_A(P)$ and $(X_2, \mathsf{m}_{(RAR)^{-1}}(S_2^{-1}))$ is a reverse invariant pair for $\mathbf{M}_A(P)$. Then with

$$\widehat{M} := \begin{bmatrix} \mathsf{m}_{A^{-1}}(S_1) & 0 \\ 0 & \mathsf{m}_{(RAR)^{-1}}(S_2^{-1}) \end{bmatrix},$$

we claim that $(\widehat{X}, \widehat{M})$ is a "generalized decomposable pair" for $\mathbf{M}_A(P)$ in the same sense as $(\widehat{X}, \widehat{S})$ is a generalized decomposable pair for P. Letting $\check{S}_1 := \mathsf{m}_{A^{-1}}(S_1)$ and $\check{S}_2 := \mathsf{m}_{(RAR)^{-1}}(S_2^{-1})$, all that remains is to see why the matrix

$$\mathcal{W}_k\big((X_1, \check{S}_1), (X_2, \check{S}_2)\big) = \begin{bmatrix} X_1 & X_2\mathsf{m}_{(RAR)^{-1}}(S_2^{-1})^{k-1} \\ X_1\mathsf{m}_{A^{-1}}(S_1) & X_2\mathsf{m}_{(RAR)^{-1}}(S_2^{-1})^{k-2} \\ \vdots & \vdots \\ X_1\mathsf{m}_{A^{-1}}(S_1)^{k-1} & X_2 \end{bmatrix} \tag{6.5}$$

has full column rank. To do this we employ a strategy analogous to the one used in the proof of Theorem 6.11.

First observe that $\mathsf{m}_A(\frac{-d}{c}) = \infty$, and hence $\mathsf{m}_{A^{-1}}(\infty) = \frac{-d}{c}$. But ∞ is not an eigenvalue of the matrix S_1, so $\frac{-d}{c}$ is not an eigenvalue of $\mathsf{m}_{A^{-1}}(S_1) = \check{S}_1$, and therefore the matrix

$(c\breve{S}_1 + dI)$ is nonsingular. Thus $\mathsf{m}_A(\breve{S}_1) = (a\breve{S}_1 + bI)(c\breve{S}_1 + dI)^{-1}$ is well defined and equal to S_1. A similar argument shows that the matrix $(b\breve{S}_2 + aI)$ is also nonsingular, and hence that $\mathsf{m}_{RAR}(\breve{S}_2) = (d\breve{S}_2 + cI)(b\breve{S}_2 + aI)^{-1} = S_2^{-1}$. Now we use these facts to develop a relationship between the matrices $\mathcal{W}_k\big((X_1, S_1), (X_2, S_2^{-1})\big)$ and $\mathcal{W}_k\big((X_1, \breve{S}_1), (X_2, \breve{S}_2)\big)$.

$$
\mathcal{W}_k\big((X_1, S_1), (X_2, S_2^{-1})\big) \cdot \begin{bmatrix} (c\breve{S}_1 + dI)^{k-1} & 0 \\ 0 & (b\breve{S}_2 + aI)^{k-1} \end{bmatrix}
$$

$$
= \mathcal{W}_k\big((X_1, \mathsf{m}_A(\breve{S}_1)), (X_2, \mathsf{m}_{RAR}(\breve{S}_2))\big) \cdot \begin{bmatrix} (c\breve{S}_1 + dI)^{k-1} & 0 \\ 0 & (b\breve{S}_2 + aI)^{k-1} \end{bmatrix}
$$

$$
= \begin{bmatrix} X_1\,(c\breve{S}_1 + dI)^{k-1} & X_2\,(d\breve{S}_2 + cI)^{k-1} \\ X_1\,(c\breve{S}_1 + dI)^{k-2}(a\breve{S}_1 + bI) & X_2\,(d\breve{S}_2 + cI)^{k-2}(b\breve{S}_2 + aI) \\ X_1\,(c\breve{S}_1 + dI)^{k-3}(a\breve{S}_1 + bI)^2 & X_2\,(d\breve{S}_2 + cI)^{k-3}(b\breve{S}_2 + aI)^2 \\ \vdots & \vdots \\ X_1\,(a\breve{S}_1 + bI)^{k-1} & X_2\,(b\breve{S}_2 + aI)^{k-1} \end{bmatrix}
$$

$$
= \begin{bmatrix} X_1\,q_1(\breve{S}_1) & X_2\,(\mathrm{rev}_{k-1}\,q_1)(\breve{S}_2) \\ X_1\,q_2(\breve{S}_1) & X_2\,(\mathrm{rev}_{k-1}\,q_2)(\breve{S}_2) \\ \vdots & \vdots \\ X_1\,q_k(\breve{S}_1) & X_2\,(\mathrm{rev}_{k-1}\,q_k)(\breve{S}_2) \end{bmatrix},
$$

where each q_i is a scalar polynomial of grade $k-1$. Assembling the coefficients of all the q_i polynomials row-wise into a $k \times k$ matrix $\breve{\mathcal{Q}}$, then the last matrix can be expressed as

$$
(\breve{\mathcal{Q}} \otimes I_n) \cdot \begin{bmatrix} X_1 & X_2\,\breve{S}_2^{k-1} \\ X_1\,\breve{S}_1 & X_2\,\breve{S}_2^{k-2} \\ \vdots & \vdots \\ X_1\,\breve{S}_1^{k-1} & X_2 \end{bmatrix} = (\breve{\mathcal{Q}} \otimes I_n) \cdot \mathcal{W}_k\big((X_1, \breve{S}_1), (X_2, \breve{S}_2)\big).
$$

Letting B be the invertible block-diagonal matrix $B = \mathrm{diag}\big((c\breve{S}_1 + dI)^{k-1}, (b\breve{S}_2 + aI)^{k-1}\big)$, we have thus shown that

$$
(\breve{\mathcal{Q}} \otimes I_n) \cdot \mathcal{W}_k\big((X_1, \breve{S}_1), (X_2, \breve{S}_2)\big) = \mathcal{W}_k\big((X_1, S_1), (X_2, S_2^{-1})\big) \cdot B. \tag{6.6}
$$

Observe that the relationship in (6.6) is the analog of that developed in (6.3) for the proof of Theorem 6.11. From (6.6) it is now immediate that

$$
\mathrm{rank}\,\mathcal{W}_k\big((X_1, \breve{S}_1), (X_2, \breve{S}_2)\big) \geq \mathrm{rank}\,\mathcal{W}_k\big((X_1, S_1), (X_2, S_2^{-1})\big),
$$

and hence that $\mathcal{W}_k\big((X_1, \breve{S}_1), (X_2, \breve{S}_2)\big)$ has full column rank, as desired. Finally, although we did not need this fact here, it is worth noting that the ranks can be shown to always be equal, by an argument analogous to the one used in the proof of Theorem 6.11.

7 Möbius and Minimal Indices

An important aspect of the structure of a singular matrix polynomial is captured by its minimal indices and bases. This section investigates the relationship between these quantities for a singular polynomial P and any of its Möbius transforms $\mathbf{M}_A(P)$. For the convenience of the reader, we begin by recalling the definition and basic properties of minimal indices and bases, and then proceed in Section 7.2 to establish the main results.

7.1 Minimal Indices and Bases – Definition and Basic Properties

Let $\mathbb{F}(\lambda)$ denote the field of rational functions over $\mathbb{F}$, and $\mathbb{F}(\lambda)^n$ the vector space of all n-vectors with entries in $\mathbb{F}(\lambda)$. In this section, elements of $\mathbb{F}(\lambda)^n$ will be denoted by notation like $\vec{x}(\lambda)$, in order to maintain a clear visual distinction between vectors in $\mathbb{F}(\lambda)^n$ and scalars in $\mathbb{F}(\lambda)$.

An $m \times n$ *singular* matrix polynomial $P(\lambda)$ has nontrivial *right* and/or *left null vectors*, that is, vectors $\vec{x}(\lambda) \in \mathbb{F}(\lambda)^n$ and $\vec{y}(\lambda) \in \mathbb{F}(\lambda)^m$ such that $P(\lambda)\vec{x}(\lambda) \equiv 0$ and $\vec{y}(\lambda)^T P(\lambda) \equiv 0$, where $\vec{y}(\lambda)^T$ denotes the transpose of $\vec{y}(\lambda)$. This leads to the following definition.

Definition 7.1. *The* right *and* left nullspaces *of the $m \times n$ matrix polynomial $P(\lambda)$, denoted respectively by $\mathcal{N}_r(P)$ and $\mathcal{N}_\ell(P)$, are the subspaces*

$$
\begin{aligned}
\mathcal{N}_r(P) &:= \left\{ \vec{x}(\lambda) \in \mathbb{F}(\lambda)^n : P(\lambda)\vec{x}(\lambda) \equiv \vec{0} \right\}, \\
\mathcal{N}_\ell(P) &:= \left\{ \vec{y}(\lambda) \in \mathbb{F}(\lambda)^m : \vec{y}(\lambda)^T P(\lambda) \equiv \vec{0}^T \right\},
\end{aligned}
$$

respectively.

A *vector polynomial* is a vector all of whose entries are polynomials in the variable λ. For any subspace of $\mathbb{F}(\lambda)^n$, it is always possible to find a basis consisting entirely of vector polynomials; simply take an arbitrary basis and multiply each vector by the denominators of its entries. The *degree* of a vector polynomial is the largest degree of its components, and the *order* of a polynomial basis is defined as the sum of the degrees of its vectors [25, p. 494]. Then we have the following definition of a minimal basis for a rational subspace.

Definition 7.2. *Let $\mathcal{S}$ be a subspace of $\mathbb{F}(\lambda)^n$. The* order *of $\mathcal{S}$, denoted* $\mathrm{ord}(\mathcal{S})$, *is the least order among all vector polynomial bases of $\mathcal{S}$, and a* minimal basis *of $\mathcal{S}$ is any vector polynomial basis of $\mathcal{S}$ that attains this least order.*

In [25] it is shown that for any given subspace $\mathcal{S}$ of $\mathbb{F}(\lambda)^n$, the ordered list of degrees of the vector polynomials in any minimal basis of $\mathcal{S}$ is always the same; these degrees are the *minimal indices* of $\mathcal{S}$. Specializing $\mathcal{S}$ to be the left and right nullspaces of a singular matrix polynomial gives Definition 7.3; here the degree of a vector polynomial $\vec{p}(\lambda)$ is denoted $\deg \vec{p}(\lambda)$.

Definition 7.3 (Minimal indices and bases of a singular matrix polynomial)**.**
For a singular matrix polynomial $P(\lambda)$, let the sets $\left\{ \vec{y}_1(\lambda), \ldots, \vec{y}_q(\lambda) \right\}$ and $\left\{ \vec{x}_1(\lambda), \ldots, \vec{x}_p(\lambda) \right\}$ be minimal bases, respectively, of the left and right nullspaces of $P(\lambda)$, ordered such that $\deg(\vec{y}_1) \leq \deg(\vec{y}_2) \leq \cdots \leq \deg(\vec{y}_q)$ and $\deg(\vec{x}_1) \leq \cdots \leq \deg(\vec{x}_p)$. Let $\eta_i = \deg(\vec{y}_i)$ for $i = 1, \ldots, q$ and $\varepsilon_j = \deg(\vec{x}_j)$ for $j = 1, \ldots, p$. Then $\eta_1 \leq \eta_2 \leq \cdots \leq \eta_q$ and $\varepsilon_1 \leq \varepsilon_2 \leq \cdots \leq \varepsilon_p$ are, respectively, the left *and* right minimal indices *of $P(\lambda)$. For the sake of brevity, we call minimal bases of the left and right nullspaces of $P(\lambda)$ simply* left *and* right minimal bases *of $P(\lambda)$.*

It is not hard to see that the minimal indices of a singular polynomial $P(\lambda)$ are invariant under *strict equivalence*, i.e., under pre- and post-multiplication of $P(\lambda)$ by nonsingular constant matrices. However, unimodular equivalence can change any (even all) of the minimal indices, as illustrated in [14, 20, 21]. Note that in the case of matrix pencils, the left (right) minimal indices can be read off from the sizes of the left (right) singular blocks of the Kronecker canonical form of the pencil [27, Chap. XII].

We refer the reader to [20, Section 2] for a more complete discussion, and to [43] for an alternative formulation that gives additional insight into the notions of minimal indices and bases.

7.2 Effect of Möbius on Minimal Indices and Bases

In this section the simple relationship between the minimal indices and bases of a matrix polynomial $P(\lambda)$ and those of any of its Möbius transforms $\mathbf{M}_A(P)$ is established. For convenience we discuss only right minimal indices and bases; the story for left minimal indices and bases is essentially the same.

We begin by observing that minimal indices and bases of a Möbius transform $\mathbf{M}_A(P)$ are independent of the choice of grade for $P(\lambda)$. Computing formally in $\mathbb{F}(\lambda)^n$, we have

$$\mathbf{M}_A(P)\,\vec{y}(\lambda) = 0 \;\Leftrightarrow\; (c\lambda + d)^k P\big(\mathsf{m}_A(\lambda)\big)\,\vec{y}(\lambda) = 0 \;\Leftrightarrow\; P\big(\mathsf{m}_A(\lambda)\big)\,\vec{y}(\lambda) = 0\,. \tag{7.1}$$

Thus the right nullspace of the matrix polynomial $\mathbf{M}_A(P)$ is identical to the right nullspace of the rational matrix $P\big(\mathsf{m}_A(\lambda)\big)$, and so it is independent of the grade k chosen for $P(\lambda)$. Since minimal indices and bases are properties of rational subspaces themselves, we also see from (7.1) that to find these quantities for $\mathbf{M}_A(P)$ it suffices to work directly with the rational matrix $P\big(\mathsf{m}_A(\lambda)\big)$. The equivalence

$$P(\lambda)\,\vec{x}(\lambda) = 0 \;\Leftrightarrow\; P\big(\mathsf{m}_A(\lambda)\big)\,\vec{x}\big(\mathsf{m}_A(\lambda)\big) = 0 \tag{7.2}$$

further suggests that the following "rational substitution" map $\mathfrak{R}_A$ will be useful in analyzing how the minimal indices of P and $\mathbf{M}_A(P)$ are related. Letting $\mathcal{V} := \mathbb{F}(\lambda)^n$ throughout this section, we define the map

$$\begin{aligned} \mathfrak{R}_A : \quad \mathcal{V} &\longrightarrow \mathcal{V} \\ \vec{v}(\lambda) &\longmapsto \vec{v}\big(\mathsf{m}_A(\lambda)\big)\,. \end{aligned} \tag{7.3}$$

If the map $\mathfrak{R}_A$ was $\mathbb{F}(\lambda)$-linear, then many of the properties listed in the next theorem would follow immediately. However, $\mathfrak{R}_A$ is *not* $\mathbb{F}(\lambda)$-linear, so we present the proof.

Theorem 7.4 (Basic properties of $\mathfrak{R}_A$).
Let $A \in GL(2,\mathbb{F})$. Then the map $\mathfrak{R}_A$ defined on the $\mathbb{F}(\lambda)$-vector space $\mathcal{V} := \mathbb{F}(\lambda)^n$ as in (7.3) enjoys the following properties:

(a) *$\mathfrak{R}_A$ is additive, i.e., $\mathfrak{R}_A(\vec{v}+\vec{w}) = \mathfrak{R}_A(\vec{v}) + \mathfrak{R}_A(\vec{w})$ for all $\vec{v},\vec{w} \in \mathcal{V}$, but is not compatible with multiplication by scalars from $\mathbb{F}(\lambda)$.*

(b) *$\mathfrak{R}_A$ is a bijection on $\mathcal{V}$ that maps subspaces to subspaces. In other words, for any $\mathbb{F}(\lambda)$-subspace $\mathcal{S} \subseteq \mathcal{V}$, the image $\mathfrak{R}_A(\mathcal{S})$ is also an $\mathbb{F}(\lambda)$-subspace of $\mathcal{V}$.*

(c) *$\mathfrak{R}_A$ preserves linear independence; i.e., if $\big\{\vec{v}_1(\lambda),\ldots,\vec{v}_\ell(\lambda)\big\}$ are linearly independent in $\mathcal{V}$, then so are $\big\{\mathfrak{R}_A(\vec{v}_1),\ldots,\mathfrak{R}_A(\vec{v}_\ell)\big\}$.*

(d) *$\mathfrak{R}_A$ preserves dimension; if $\mathcal{S} \subseteq \mathcal{V}$ is any subspace, then $\dim_{\mathbb{F}(\lambda)} \mathfrak{R}_A(\mathcal{S}) = \dim_{\mathbb{F}(\lambda)} \mathcal{S}$. Furthermore, $\big\{\vec{v}_1(\lambda),\ldots,\vec{v}_k(\lambda)\big\}$ is a basis for $\mathcal{S}$ if and only if $\big\{\mathfrak{R}_A(\vec{v}_1),\ldots,\mathfrak{R}_A(\vec{v}_k)\big\}$ is a basis for $\mathfrak{R}_A(\mathcal{S})$.*

(e) *Suppose $\big\{\vec{v}_1(\lambda),\ldots,\vec{v}_k(\lambda)\big\}$ is any vector polynomial basis for a rational subspace $\mathcal{S} \subseteq \mathcal{V}$. Then $\big\{\mathbf{M}_A(\vec{v}_1),\ldots,\mathbf{M}_A(\vec{v}_k)\big\}$, where each $\mathbf{M}_A(\vec{v}_j)$ is taken with respect to the degree of $\vec{v}_j$, is a vector polynomial basis for $\mathfrak{R}_A(\mathcal{S})$, with $\deg \mathbf{M}_A(\vec{v}_j) \leq \deg \vec{v}_j$ for $j = 1,\ldots,k$.*

(f) *$\mathfrak{R}_A$ preserves the order and the minimal indices of rational subspaces; if $\mathcal{S} \subseteq \mathcal{V}$ is any subspace, then $\mathrm{ord}\big(\mathfrak{R}_A(\mathcal{S})\big) = \mathrm{ord}(\mathcal{S})$, and the minimal indices of the subspaces $\mathfrak{R}_A(\mathcal{S})$ and $\mathcal{S}$ are identical. Indeed, if $\big\{\vec{v}_1(\lambda),\ldots,\vec{v}_k(\lambda)\big\}$ is a minimal basis for the subspace $\mathcal{S}$, then $\big\{\mathbf{M}_A(\vec{v}_1),\ldots,\mathbf{M}_A(\vec{v}_k)\big\}$ is a minimal basis for the subspace $\mathfrak{R}_A(\mathcal{S})$; here as in part (e) each Möbius transform $\mathbf{M}_A(\vec{v}_j)$ is taken with respect to the degree of $\vec{v}_j$.*

Proof. (a) The additive property of $\mathfrak{R}_A$ is immediate from the definition. For scalar multiplication, clearly $\mathfrak{R}_A\big(\alpha(\lambda)\vec{v}(\lambda)\big) = \mathfrak{R}_A\big(\alpha(\lambda)\big)\mathfrak{R}_A\big(\vec{v}(\lambda)\big)$ always holds, but this will be equal to $\alpha(\lambda)\mathfrak{R}_A\big(\vec{v}(\lambda)\big)$ only if $\mathfrak{R}_A\big(\alpha(\lambda)\big) = \alpha\big(\mathsf{m}_A(\lambda)\big)$ and $\alpha(\lambda)$ are equal as elements of $\mathbb{F}(\lambda)$, which will not be true in general. Thus, in general, $\mathfrak{R}_A$ is incompatible with $\mathbb{F}(\lambda)$-scalar multiplication.

(b) That $\mathfrak{R}_A$ is a bijection follows from the existence of an inverse map, which is easily seen to be just $\mathfrak{R}_{A^{-1}}$. To see that $\mathfrak{R}_A(\mathcal{S})$ is a subspace whenever $\mathcal{S} \subseteq \mathcal{V}$ is, we directly verify the two closure conditions. First suppose that $\vec{v}, \vec{w} \in \mathfrak{R}_A(\mathcal{S})$, so that $\vec{v}(\lambda) = \mathfrak{R}_A\big(\vec{x}(\lambda)\big)$ and $\vec{w}(\lambda) = \mathfrak{R}_A\big(\vec{y}(\lambda)\big)$ for some $\vec{x}, \vec{y} \in \mathcal{S}$. Then $\vec{v} + \vec{w} = \mathfrak{R}_A(\vec{x} + \vec{y})$ by the additivity of $\mathfrak{R}_A$ from part (a), and so the additive closure of $\mathfrak{R}_A(\mathcal{S})$ follows from the additive closure of $\mathcal{S}$.

To verify that $\mathfrak{R}_A(\mathcal{S})$ is closed under multiplication by scalars from the field $\mathbb{F}(\lambda)$, consider an arbitrary scalar $\alpha(\lambda) \in \mathbb{F}(\lambda)$ and vector $\vec{v}(\lambda) \in \mathfrak{R}_A(\mathcal{S})$. Let $\vec{x} \in \mathcal{S}$ be such that $\vec{v} = \mathfrak{R}_A(\vec{x})$, and let $\beta(\lambda) := \alpha\big(\mathsf{m}_{A^{-1}}(\lambda)\big) \in \mathbb{F}(\lambda)$. Then $\beta(\lambda)\vec{x}(\lambda) \in \mathcal{S}$ since $\mathcal{S}$ is an $\mathbb{F}(\lambda)$-subspace, and it is easily seen that $\alpha(\lambda)\vec{v}(\lambda) = \alpha(\lambda)\mathfrak{R}_A\big(\vec{x}(\lambda)\big) = \mathfrak{R}_A\big(\beta(\lambda)\vec{x}(\lambda)\big) \in \mathfrak{R}_A(\mathcal{S})$. Thus $\mathfrak{R}_A(\mathcal{S})$ is closed under scalar multiplication, and hence is an $\mathbb{F}(\lambda)$-subspace.

(c) Suppose $\sum_{j=1}^{\ell} \alpha_j(\lambda)\mathfrak{R}_A(\vec{v}_j) = 0$. Then by part (a) we have

$$\mathfrak{R}_{A^{-1}}\left(\sum_{j=1}^{\ell} \alpha_j(\lambda)\mathfrak{R}_A(\vec{v}_j)\right) = \sum_{j=1}^{\ell} \mathfrak{R}_{A^{-1}}\big(\alpha_j(\lambda)\big) \cdot \vec{v}_j = 0\,,$$

and so $\mathfrak{R}_{A^{-1}}\big(\alpha_j(\lambda)\big) = 0$ for each $j = 1, \ldots, \ell$, since $\big\{\vec{v}_1(\lambda), \ldots, \vec{v}_\ell(\lambda)\big\}$ are linearly independent. But by part (b) for the case $n = 1$, $\mathfrak{R}_{A^{-1}}$ is a bijection, so $\alpha_j(\lambda) = 0$ for $j = 1, \ldots, \ell$. Thus $\big\{\mathfrak{R}_A(\vec{v}_1), \ldots, \mathfrak{R}_A(\vec{v}_\ell)\big\}$ are also linearly independent.

(d) If $\big\{\vec{v}_1(\lambda), \ldots, \vec{v}_k(\lambda)\big\}$ is a basis for $\mathcal{S}$, then $\big\{\mathfrak{R}_A(\vec{v}_1), \ldots, \mathfrak{R}_A(\vec{v}_k)\big\}$ is linearly independent in $\mathfrak{R}_A(\mathcal{S})$ by part (c), so

$$\dim \mathfrak{R}_A(\mathcal{S}) \geq k = \dim \mathcal{S}\,.$$

Exchanging the roles of $\mathcal{S}$ and $\mathfrak{R}_A(\mathcal{S})$, and replacing $\mathfrak{R}_A$ by $\mathfrak{R}_{A^{-1}}$, the same argument shows that $\dim \mathcal{S} \geq \dim \mathfrak{R}_A(\mathcal{S})$. Hence $\dim \mathfrak{R}_A(\mathcal{S}) = \dim \mathcal{S}$, and $\big\{\vec{v}_1(\lambda), \ldots, \vec{v}_k(\lambda)\big\}$ is a basis for $\mathcal{S}$ if and only if $\big\{\mathfrak{R}_A(\vec{v}_1), \ldots, \mathfrak{R}_A(\vec{v}_k)\big\}$ is a basis for $\mathfrak{R}_A(\mathcal{S})$.

(e) If $\big\{\vec{v}_1(\lambda), \ldots, \vec{v}_k(\lambda)\big\}$ is a *vector polynomial* basis for $\mathcal{S}$, then certainly each $\mathbf{M}_A(\vec{v}_i)$ is a vector polynomial by definition of $\mathbf{M}_A$. Choosing the grade of each $\vec{v}_i$ to be equal to $\deg \vec{v}_i$, we then also have $\deg \mathbf{M}_A(\vec{v}_i) \leq \operatorname{grade} \mathbf{M}_A(\vec{v}_i) = \deg \vec{v}_i$. By part (d) we know that $\big\{\mathfrak{R}_A(\vec{v}_1), \ldots, \mathfrak{R}_A(\vec{v}_k)\big\}$ is a *rational* basis for $\mathfrak{R}_A(\mathcal{S})$. But each $\mathbf{M}_A(\vec{v}_i)$ is just a nonzero $\mathbb{F}(\lambda)$-scalar multiple of $\mathfrak{R}_A(\vec{v}_i)$, so $\big\{\mathbf{M}_A(\vec{v}_1), \ldots, \mathbf{M}_A(\vec{v}_k)\big\}$ must also be a basis for $\mathfrak{R}_A(\mathcal{S})$.

(f) Suppose $\big\{\vec{v}_1(\lambda), \ldots, \vec{v}_k(\lambda)\big\}$ is a *minimal basis* for $\mathcal{S}$. Then using part (e) we can see that $\operatorname{ord}\big(\mathfrak{R}_A(\mathcal{S})\big) \leq \operatorname{ord}(\mathcal{S})$, since

$$\operatorname{ord}\big(\mathfrak{R}_A(\mathcal{S})\big) \leq \sum_{j=1}^{k} \deg \mathbf{M}_A(\vec{v}_j) \leq \sum_{j=1}^{k} \deg \vec{v}_j(\lambda) = \operatorname{ord}(\mathcal{S})\,. \tag{7.4}$$

Applying this result to $\mathfrak{R}_A(\mathcal{S})$ in place of $\mathcal{S}$, and to the rational substitution map $\mathfrak{R}_{A^{-1}}$, shows that

$$\operatorname{ord}(\mathcal{S}) = \operatorname{ord}\Big(\mathfrak{R}_{A^{-1}}\big(\mathfrak{R}_A(\mathcal{S})\big)\Big) \leq \operatorname{ord}\big(\mathfrak{R}_A(\mathcal{S})\big)\,,$$

and so we have $\operatorname{ord}\big(\mathfrak{R}_A(\mathcal{S})\big) = \operatorname{ord}(\mathcal{S})$, as desired. Furthermore we must have equalities everywhere in (7.4). Since $\deg \mathbf{M}_A(v_j) \leq \deg(v_j)$ for *each* $j = 1, \ldots, k$ by part (e), we can only have equality in (7.4) if $\deg \mathbf{M}_A(v_j) = \deg(v_j)$ for every $j = 1, \ldots, k$. Therefore the minimal indices of $\mathfrak{R}_A(\mathcal{S})$ must be identical to those of $\mathcal{S}$, and $\big\{\mathbf{M}_A(\vec{v}_1), \ldots, \mathbf{M}_A(\vec{v}_k)\big\}$ must be a minimal basis for $\mathfrak{R}_A(\mathcal{S})$. $\square$

From Theorem 7.4 we can now easily obtain the main result of this section, which in brief says that any Möbius transformation preserves the *singular structure* of a singular polynomial.

Theorem 7.5 (Möbius transformations preserve minimal indices)**.**
If P is a singular matrix polynomial over $\mathbb{F}$ and $A \in GL(2, \mathbb{F})$, then the minimal indices of P and $\mathbf{M}_A(P)$ are identical. Furthermore, if $\{\vec{x}_1(\lambda), \ldots, \vec{x}_p(\lambda)\}$ and $\{\vec{y}_1(\lambda), \ldots, \vec{y}_q(\lambda)\}$ are right and left minimal bases for P, then $\{\mathbf{M}_A(\vec{x}_1), \ldots, \mathbf{M}_A(\vec{x}_p)\}$ and $\{\mathbf{M}_A(\vec{y}_1), \ldots, \mathbf{M}_A(\vec{y}_q)\}$ are right and left minimal bases, respectively, for $\mathbf{M}_A(P)$, where each Möbius transform $\mathbf{M}_A(\vec{v})$ of a vector polynomial $\vec{v}$ is taken with respect to the degree of $\vec{v}$.

Proof. We consider only the right minimal indices and bases; the argument for the left minimal indices and bases is entirely analogous. The key step is to show that the right nullspace of $\mathbf{M}_A(P)$ can be characterized as

$$\mathcal{N}_r\big(\mathbf{M}_A(P)\big) \;=\; \mathfrak{R}_A\big(\mathcal{N}_r(P)\big). \tag{7.5}$$

Then applying Theorem 7.4(f) to the subspace $\mathcal{S} = \mathcal{N}_r(P)$ will imply the desired results.

First observe that we have already proved a substantial part of (7.5). From (7.1) we can immediately conclude that $\mathcal{N}_r\big(\mathbf{M}_A(P)\big) = \mathcal{N}_r\big(\mathfrak{R}_A(P)\big)$, and (7.2) is equivalent to the containment $\mathfrak{R}_A\big(\mathcal{N}_r(P)\big) \subseteq \mathcal{N}_r\big(\mathfrak{R}_A(P)\big)$. All that remains to complete the proof of (7.5), and hence of the theorem, is to establish the reverse containment $\mathcal{N}_r\big(\mathfrak{R}_A(P)\big) \subseteq \mathfrak{R}_A\big(\mathcal{N}_r(P)\big)$.

So suppose that $\vec{y} \in \mathcal{N}_r\big(\mathfrak{R}_A(P)\big)$, i.e., that $P\big(\mathsf{m}_A(\lambda)\big) \cdot \vec{y}(\lambda) = 0$. Substituting $\mathsf{m}_{A^{-1}}(\lambda)$ for λ gives

$$0 \;=\; P\big(\mathsf{m}_A(\mathsf{m}_{A^{-1}}(\lambda))\big) \cdot \vec{y}\big(\mathsf{m}_{A^{-1}}(\lambda)\big) \;=\; P(\lambda) \cdot \vec{x}(\lambda) \,,$$

where $\vec{x}(\lambda) := \vec{y}\big(\mathsf{m}_{A^{-1}}(\lambda)\big)$. Clearly $\vec{x}(\lambda) \in \mathcal{N}_r(P)$ and $\vec{y}(\lambda) = \mathfrak{R}_A\big(\vec{x}(\lambda)\big)$, so $\vec{y}(\lambda) \in \mathfrak{R}_A\big(\mathcal{N}_r(P)\big)$, and the containment $\mathcal{N}_r\big(\mathfrak{R}_A(P)\big) \subseteq \mathfrak{R}_A\big(\mathcal{N}_r(P)\big)$ is established, thus completing the proof. $\qquad\square$

Remark 7.6. As a final observation, it is interesting to note that the results in this section provide an alternative (and independent) proof of the rank preservation property of Möbius transformations, proved earlier in Proposition 3.29. Since the dimensions of $\mathcal{N}_r(P)$ and $\mathcal{N}_r\big(\mathbf{M}_A(P)\big)$ are equal by equation (7.5) and Theorem 7.4(d), or by Theorem 7.5 itself, the ranks of P and $\mathbf{M}_A(P)$ must be the same by the rank-nullity theorem.

8 Möbius and Linearizations

In this section we investigate the effect of Möbius transformations on matrix polynomials and their linearizations. In particular, our aim is to characterize those Möbius transformations that preserve linearizations; that is, if a pencil L is a linearization for a matrix polynomial P, then when is $\mathbf{M}_A(L)$ a linearization for $\mathbf{M}_A(P)$? To aid in this investigation we begin by recalling two recently introduced equivalence relations and their properties.

8.1 Spectral Equivalence of Matrix Polynomials

Recently the relations of extended unimodular equivalence and spectral equivalence have been introduced [23] to facilitate the comparison of matrix polynomials of different sizes and/or grades. The underlying goal in [23] is to investigate the extent to which it is possible for matrix polynomials of different grades to have the same elementary divisors *and* the same minimal indices. Motivating this investigation is the classical technique for solving the eigenproblem for a polynomial $P(\lambda)$ by the use of a linearization, i.e., a matrix pencil with the same elementary divisors as $P(\lambda)$. Taking the standard definitions of linearization and strong linearization [29, 30, 41, 45] as prototypes leads to the following.

Definition 8.1 (Extended unimodular equivalence and spectral equivalence)**.**
Consider matrix polynomials P and Q, of any sizes and grades, over an arbitrary field $\mathbb{F}$.

(a) *P and Q are said to be* extended unimodular equivalent, *denoted $P \smile Q$, if there exist some $r, s \in \mathbb{N}$ such that $\operatorname{diag}[P, I_r] \sim \operatorname{diag}[Q, I_s]$, i.e., such that $\operatorname{diag}[P, I_r]$ and $\operatorname{diag}[Q, I_s]$ are unimodularly equivalent over $\mathbb{F}$.*

(b) *P and Q are said to be* spectrally equivalent, *denoted $P \asymp Q$, if both $P \smile Q$ and $\mathbf{M}_R(P) \smile \mathbf{M}_R(Q)$.*

(Recall that $\mathbf{M}_R(P)$ denotes reversal with respect to the grade of the polynomial P, as described in Example 3.7.)

It is not hard to see that Definitions 8.1(a) and (b) do indeed define equivalence relations. Also note two key features of these definitions: the flexibility in what sizes are allowed for equivalent polynomials, as well as the almost complete absence of any mention of degree or grade. Indeed, grade and degree play no role at all in the definition of extended unimodular equivalence, and appear only implicitly in the use of reversals in the definition of spectral equivalence.

Justification for the name "spectral equivalence" is provided by the following theorem, which characterizes each of these equivalence relations in terms of spectral data.

Theorem 8.2 (Characterization of spectral equivalence [23])**.**
Let P and Q be any pair of matrix polynomials over a field $\mathbb{F}$, and let $\overline{\mathbb{F}}$ denote the algebraic closure of $\mathbb{F}$. Consider the following properties:

(a) *$\widehat{\mathcal{J}}(P, \lambda_0) = \widehat{\mathcal{J}}(Q, \lambda_0)$ for every $\lambda_0 \in \overline{\mathbb{F}}$, i.e., P and Q have the same (finite) elementary divisors over $\overline{\mathbb{F}}$,*

(b) *$\widehat{\mathcal{J}}(P, \infty) = \widehat{\mathcal{J}}(Q, \infty)$, i.e., P and Q have the same infinite elementary divisors,*

(c) *the dimensions of corresponding nullspaces are equal, that is,*
$$\dim \mathcal{N}_\ell(P) = \dim \mathcal{N}_\ell(Q) \quad and \quad \dim \mathcal{N}_r(P) = \dim \mathcal{N}_r(Q).$$

Then $P \smile Q$ if and only if (a) and (c) hold, while $P \asymp Q$ if and only if (a), (b), and (c) hold.

Note that condition (c) of Theorem 8.2 is equivalent to saying that P and Q have the same *number* of left minimal indices, and the same *number* of right minimal indices. Thus for two matrix polynomials to have any chance of having the same elementary divisors *and* minimal indices, Theorem 8.2 makes it clear that a necessary first step is that they be spectrally equivalent.

8.2 Linearizations and Quadratifications

The classical notions of linearization and strong linearization were used as models for the definitions of extended unimodular equivalence and spectral equivalence. We now come full circle, using these equivalence relations to extend the classical notions of linearization to concepts that are more flexible and easier to use.

Definition 8.3 (Linearizations and Quadratifications [23])**.**
Let P be a matrix polynomial of any size and grade, over an arbitrary field $\mathbb{F}$.

(a) *A* linearization *for P is a pencil L (i.e., a polynomial of grade one) such that $L \smile P$, while a* strong linearization *is a pencil L such that $L \asymp P$.*

(b) *A* quadratification *for P is a quadratic polynomial Q (i.e., a polynomial of grade two) such that $Q \smile P$, while a* strong quadratification *is a quadratic polynomial Q such that $Q \asymp P$.*

Clearly any linearization or strong linearization in the classical sense is still one according to this definition. By contrast, linearizations and quadratifications in the sense of Definition 8.3 (strong or not) are no longer restricted a priori to one fixed size, but are free to take on any size that "works", i.e., that makes $L \smile P$, or $L \asymp P$, or $Q \smile P$, or $Q \asymp P$. This is in accord with the known situation for rectangular matrix polynomials, where there always exist linearizations of many possible sizes [22], and with the recent development of the notion of "trimmed" linearizations [14].

However, for regular polynomials P it turns out that only one size, the classical size, can actually occur for strong linearizations in this new sense, even though Definition 8.3 allows the a priori possibility of other sizes.

Theorem 8.4 (Size of strong linearizations for regular polynomials [23]).
Suppose P is a regular matrix polynomial of size $n \times n$ and grade k. Then any strong linearization for P must have the classical size $kn \times kn$.

Thus in the regular case the new definition is equivalent to the classical one, and so in a certain sense can be viewed as a conservative extension of the classical definition. The situation for singular polynomials is rather different, however; in this case there are always many possible sizes for linearizations and strong linearizations in the sense of Definition 8.3 [23].

8.3   Möbius and Spectral Equivalence

We now present the main results of Section 8, showing that every Möbius transformation preserves the relation of spectral equivalence, and consequently that every strong linearization is preserved by Möbius transformations. Finally we consider the effect of Möbius transformations on linearizations that are not strong.

Theorem 8.5 (Möbius preserves spectral equivalence).
Let P and Q be any two matrix polynomials over an arbitrary field $\mathbb{F}$, and $A \in GL(2, \mathbb{F})$. Then $P \asymp Q$ if and only if $\mathbf{M}_A(P) \asymp \mathbf{M}_A(Q)$.

Proof. ($\Rightarrow$):   Suppose that $P \asymp Q$. Since $\mathbf{M}_A$ preserves size of matrix polynomials by definition, and rank by Proposition 3.29, we see that $\dim \mathcal{N}_r\big(\mathbf{M}_A(P)\big) = \dim \mathcal{N}_r(P)$ by the rank-nullity theorem. Thus we have

$$\dim \mathcal{N}_r\big(\mathbf{M}_A(P)\big) = \dim \mathcal{N}_r(P) = \dim \mathcal{N}_r(Q) = \dim \mathcal{N}_r\big(\mathbf{M}_A(Q)\big),$$

where the middle equality is by condition (c) in Theorem 8.2. A similar argument shows that $\dim \mathcal{N}_\ell\big(\mathbf{M}_A(P)\big) = \dim \mathcal{N}_\ell\big(\mathbf{M}_A(Q)\big)$.

On the other hand, conditions (a) and (b) in Theorem 8.2 imply that P and Q have the same elementary divisors. But from Theorem 5.3 and Remark 5.4 we know that $\mathbf{M}_A$ transforms the elementary divisors of P and Q in the same way, so that $\mathbf{M}_A(P)$ and $\mathbf{M}_A(Q)$ must also have the same elementary divisors. Thus by Theorem 8.2 we conclude that $\mathbf{M}_A(P) \asymp \mathbf{M}_A(Q)$.

($\Leftarrow$):   Apply ($\Rightarrow$) to the relation $\mathbf{M}_A(P) \asymp \mathbf{M}_A(Q)$ using the Möbius transformation $\mathbf{M}_{A^{-1}}$, to conclude that $P \asymp Q$. $\qquad\square$

Corollary 8.6 (Möbius preserves strong linearizations and quadratifications).
Suppose P is any matrix polynomial (regular or singular), over an arbitrary field $\mathbb{F}$, and $A \in GL(2, \mathbb{F})$. Then:

(a) *L is a strong linearization for P if and only if $\mathbf{M}_A(L)$ is a strong linearization for $\mathbf{M}_A(P)$.*

(b) Q is a strong quadratification for P if and only if $\mathbf{M}_A(Q)$ is a strong quadratification for $\mathbf{M}_A(P)$.

Proof. Both (a) and (b) are just special cases of Theorem 8.5. □

From Corollary 8.6 we now know that any strong linearization is preserved by every Möbius transformation. But what about linearizations that are not strong, sometimes referred to as "weak" linearizations [41]? When are they preserved by Möbius transformations? The following result completely answers this question, by characterizing the Möbius transformations that preserve any given weak linearization.

Theorem 8.7 (Möbius and weak linearizations)**.**
Let P be a matrix polynomial over an arbitrary field $\mathbb{F}$, and $A \in GL(2, \mathbb{F})$. Suppose L is a linearization for P that is not strong, i.e., $L \smallsmile P$ but $L \not\approx P$. Then $\mathbf{M}_A(L)$ is a linearization for $\mathbf{M}_A(P)$ if and only if A is upper triangular, that is, if and only if the associated Möbius function $\mathsf{m}_A(\lambda)$ can be expressed in the form $\mathsf{m}_A(\lambda) = a\lambda + b$.

Proof. To see when $\mathbf{M}_A(L)$ is a linearization for $\mathbf{M}_A(P)$, we determine when conditions (a) and (c) of Theorem 8.2 are satisfied. From Theorem 7.5 we see that condition (c) for L and P immediately implies that condition (c) also holds for $\mathbf{M}_A(L)$ and $\mathbf{M}_A(P)$, for *any* Möbius transformation $\mathbf{M}_A$. It only remains, then, to consider the elementary divisor conditions in Theorem 8.2.

For L to be a linearization that is not strong means that condition (a) in Theorem 8.2 holds, but condition (b) does not, i.e., that $\widehat{\mathcal{J}}(L, \infty) \neq \widehat{\mathcal{J}}(P, \infty)$. This implies that the chain of equalities

$$\widehat{\mathcal{J}}\big(\mathbf{M}_A(L), \lambda\big) \;=\; \widehat{\mathcal{J}}\big(L, \mathsf{m}_A(\lambda)\big) \;=\; \widehat{\mathcal{J}}\big(P, \mathsf{m}_A(\lambda)\big) \;=\; \widehat{\mathcal{J}}\big(\mathbf{M}_A(P), \lambda\big)$$

will hold at all λ with a single exception, the unique $\lambda_0 \in \overline{\mathbb{F}}_\infty$ such that $\mathsf{m}_A(\lambda_0) = \infty$. Thus condition (a) will hold for $\mathbf{M}_A(L)$ and $\mathbf{M}_A(P)$ if and only if $\lambda_0 = \infty$. But a Möbius function m_A has the property $\mathsf{m}_A(\infty) = \infty$ if and only if A is upper triangular, so the theorem is proved. □

9 Möbius and Structure

The interaction of Möbius transformations with various aspects of matrix polynomial structure has been a central theme throughout this paper. We have seen that many structural features and relationships are preserved by any Möbius transformation, e.g.:

- regularity, singularity, rank

- symmetry and skew-symmetry

- minimal indices

- the relations of strict equivalence $\approx$, and spectral equivalence $\asymp$, i.e.,

$$P \approx Q \;\Leftrightarrow\; \mathbf{M}_A(P) \approx \mathbf{M}_A(Q), \quad \text{and} \quad P \asymp Q \;\Leftrightarrow\; \mathbf{M}_A(P) \asymp \mathbf{M}_A(Q)$$

- the property of being a strong linearization or strong quadratification.

A variety of other structural features are not preserved by Möbius transformations, but instead change in a simple and predictable way. Examples of this include:

- determinants and compound matrices

- Jordan characteristic and Smith form

- invariant pairs

- minimal bases.

We now illustrate the usefulness of Möbius transformations with two examples that focus on structured matrix polynomials. In Section 9.1 we consider the problem of realizing a given list of elementary divisors by a regular matrix polynomial, in particular by a regular polynomial that is *upper triangular*; here the emphasis is on the preservation of yet another kind of structure by Möbius transformations – sparsity patterns. Next, in Section 9.2 we revisit the Cayley transformation, one of the primary examples originally motivating the work in this paper. Using the basic properties of general Möbius transformations established in Section 3, the connection between palindromic and alternating matrix polynomial structure, first established in [44], will now be made transparent. Finally, the parallelism between the Smith forms and elementary divisor structures of alternating and palindromic polynomials [46, 47] will be seen as a special case of the results in Section 5.

9.1 Sparsity and a Realization Theorem

Recall the observation made in Section 3.3 (immediately after Example 3.14), that Möbius transformations act *entry-wise* on matrix polynomials. An immediate consequence is that Möbius transformations preserve sparsity patterns in the following sense.

Remark 9.1. Let $P(\lambda)$ be a matrix polynomial of grade k, let $i, j \in \mathbb{N}$ be fixed, and let $A \in GL(2, \mathbb{F})$. If the (i, j)-entry of $P(\lambda)$, viewed as a polynomial matrix, is zero, then the (i, j)-entry of $\mathbf{M}_A(P)(\lambda)$ is also zero. In particular:

- $\mathbf{M}_A(P)(\lambda)$ is diagonal if and only if $P(\lambda)$ is diagonal.

- $\mathbf{M}_A(P)(\lambda)$ is upper triangular if and only if $P(\lambda)$ is upper triangular.

This seemingly minor observation has an important impact on inverse polynomial eigenvalue problems. In [62], the authors ask whether any regular quadratic matrix polynomial can be transformed to an upper triangular quadratic matrix polynomial having the same finite and infinite elementary divisors. A partial answer to this problem was implicit in the proof of [30, Theorem 1.7], providing the solution to an inverse problem for linearizations. The argument used in [30] actually proves a stronger result that is valid over any algebraically closed field. This stronger result appears explicitly in [61, Lemma 3.2]; we restate it here in slightly modified form.

Lemma 9.2 (Realization lemma).
Let $\mathbb{F}$ be an algebraically closed field, and let $D(\lambda)$ be a regular $n \times n$ matrix polynomial over $\mathbb{F}$ that is in Smith form. Suppose $\deg\big(\det D(\lambda)\big) = nk$ for some $k \in \mathbb{N}$. Then there exists an $n \times n$ upper triangular matrix polynomial $P(\lambda)$ of degree k, with nonsingular leading coefficient, such that $P(\lambda) := E(\lambda)D(\lambda)F(\lambda)$ where $E(\lambda)$ and $F(\lambda)$ are unimodular $n \times n$ matrix polynomials.

For the sake of completeness and for the convenience of the reader, we establish a slightly stronger result in Lemma 9.3. Setting $m = 0$ gives Lemma 9.2 as a special case.

Lemma 9.3. *Let $\mathbb{F}$ be an algebraically closed field, and let $D(\lambda) = \mathrm{diag}(d_1(\lambda), \ldots, d_n(\lambda))$ be a regular $n \times n$ matrix polynomial over $\mathbb{F}$ that is in Smith form. Suppose $\deg\big(\det D(\lambda)\big) = nk$ for some $k \in \mathbb{N}$. Further, let $B(\lambda) = \big[b_{ij}(\lambda)\big]$ be any $n \times m$ matrix polynomial with $\deg b_{ij}(\lambda) \leq \min\big(k, \deg d_i(\lambda)\big)$ for all $j = 1, \ldots, m$, $i = 1, \ldots, n$. Then there exists an upper triangular matrix polynomial $P(\lambda)$ of degree k, with nonsingular leading coefficient, such that $P(\lambda) := E(\lambda)D(\lambda)F(\lambda)$, where $E(\lambda)$ and $F(\lambda)$ are unimodular $n \times n$ matrix polynomials such that the degree of $E(\lambda)B(\lambda)$ does not exceed k.*

Proof. We proceed by induction on n, using the construction from [30, Theorem 1.7].

Assume that $n > 1$, as there is nothing to prove when $n = 1$. For convenience, let us abbreviate $\deg d_i(\lambda) =: k_i$. Since $D(\lambda)$ is in Smith form, we have $0 \leq k_1 \leq \cdots \leq k_n$, and by hypothesis

$$\sum_{i=1}^{n} k_i = \deg\big(\det D(\lambda)\big) = nk.$$

If all k_i are equal, then they are all equal to k and we are done. So suppose that not all k_i are equal. Then $k_1 < k$ and $k < k_n$, and hence $k_1 < k_1 + (k_n - k) < k_n$. Therefore there exists an index $j \geq 2$ such that $k_{j-1} \leq k_1 + (k_n - k) < k_j$, or equivalently,

$$k_{j-1} - k_1 \leq k_n - k < k_j - k_1. \tag{9.1}$$

Our first goal is to show that there exists a monic scalar polynomial $p(\lambda)$ of degree $k_n - k$ such that $d_{j-1}(\lambda)$ divides $d_1(\lambda)p(\lambda)$, and $d_1(\lambda)p(\lambda)$ divides $d_j(\lambda)$. Since $D(\lambda)$ is in Smith form, $d_1(\lambda)$ divides $d_{j-1}(\lambda)$ and $d_{j-1}(\lambda)$ divides $d_j(\lambda)$, i.e., there exist monic scalar polynomials $p_1(\lambda)$ and $p_2(\lambda)$ such that

$$d_{j-1}(\lambda) = d_1(\lambda)p_1(\lambda) \quad \text{and} \quad d_j(\lambda) = d_1(\lambda)p_1(\lambda)p_2(\lambda).$$

Now observe that $p_1(\lambda)$ has degree $k_{j-1} - k_1$ and $p_1(\lambda)p_2(\lambda)$ has degree $k_j - k_1$; also observe that $p_2(\lambda)$ decomposes into a product of linear factors, since $\mathbb{F}$ is algebraically closed. Thus by (9.1) we can remove linear factors of $p_2(\lambda)$ in the product $p_1(\lambda)p_2(\lambda)$, until the remaining polynomial $p(\lambda)$ has degree $k_n - k$; this $p(\lambda)$ has the desired properties. Now since $d_1(\lambda)p(\lambda)$ also divides $d_n(\lambda)$, let $q(\lambda)$ be a monic polynomial such that $d_1(\lambda)p(\lambda)q(\lambda) = -d_n(\lambda)$. Then $d_1(\lambda)q(\lambda)$ has degree k.

Next, perform the following elementary operations on the matrix polynomial $\big[D(\lambda), B(\lambda)\big]$:

(1) add the first column multiplied by $p(\lambda)$ to the nth column;

(2) add the first row multiplied by $q(\lambda)$ to the last row;

(3) interchange the first and nth columns;

(4) permute the rows such that row $i + 1$ goes to i for $i = 1, \ldots, j - 2$ and row one goes to row $j - 1$;

(5) permute the columns such that column $i + 1$ goes to i for $i = 1, \ldots, j - 2$ and column one goes to column $j - 1$.

44

Then the resulting matrix polynomial $\widetilde{P}(\lambda)$ has the form

$$\widetilde{P}(\lambda) \;\; = \;\; \begin{bmatrix} \widetilde{D}(\lambda) & \hat{D}(\lambda) & B_1(\lambda) \\ 0 & d_1(\lambda)q(\lambda) & B_2(\lambda) \end{bmatrix}$$

$$:= \left[\begin{array}{ccccccc|c|ccc} d_2 & & & & & & & & b_{21} & \ldots & b_{2m} \\ & \ddots & & & & & & & \vdots & \ddots & \vdots \\ & & d_{j-1} & & & & & & b_{j-1,1} & \ldots & b_{j-1,m} \\ & & & d_1 p & & & & d_1 & b_{11} & \ldots & b_{1m} \\ & & & & d_j & & & & b_{j1} & \ldots & b_{jm} \\ & & & & & \ddots & & & \vdots & \ddots & \vdots \\ & & & & & & d_{n-1} & & b_{n-1,1} & \ldots & b_{n-1,m} \\ \hline & & & & & & & d_1 q & b_{n1}+b_{11}q & \ldots & b_{nm}+b_{1m}q \end{array}\right],$$

where $\widetilde{D}(\lambda)$ is $(n-1)\times(n-1)$ and where for convenience we have suppressed the dependency on λ in the entries. Observe that only elementary row operations have been performed on $B(\lambda)$, so the transformations performed on the polynomial $\begin{bmatrix} D(\lambda), B(\lambda) \end{bmatrix}$ result in the existence of unimodular matrix polynomials $\widehat{E}(\lambda)$ and $\widehat{F}(\lambda)$ such that

$$\widehat{E}(\lambda)D(\lambda)\widehat{F}(\lambda) = \begin{bmatrix} \widetilde{D}(\lambda) & \widehat{D}(\lambda) \\ 0 & d_1(\lambda)q(\lambda) \end{bmatrix} \quad \text{and} \quad \widehat{E}(\lambda)B(\lambda) = \begin{bmatrix} B_1(\lambda) \\ B_2(\lambda) \end{bmatrix}.$$

Since $\deg b_{1j}(\lambda) \le \deg d_1(\lambda)$ for $j = 1, \ldots, m$, and since the degree of each entry of $B(\lambda)$ does not exceed k, it follows that the degree of $B_2(\lambda)$ does not exceed k either. By construction, $d_{j-1}(\lambda)$ divides $d_1(\lambda)p(\lambda)$ which in turn divides $d_j(\lambda)$, so $\widetilde{D}(\lambda)$ is in Smith form. Moreover, $d_1(\lambda)q(\lambda)$ has degree k which implies that $\deg\big(\det\widetilde{D}(\lambda)\big) = (n-1)k$. Observe that $\widetilde{D}(\lambda)$ and the $(n-1)\times(m+1)$ matrix polynomial $\widetilde{B}(\lambda) = \begin{bmatrix} \widehat{D}(\lambda), B_1(\lambda) \end{bmatrix}$ satisfy the hypotheses of the claim, so by the induction hypothesis there exist unimodular matrix polynomials $\widetilde{E}(\lambda)$ and $\widetilde{F}(\lambda)$ of size $(n-1)\times(n-1)$ such that $\widetilde{E}(\lambda)\widetilde{D}(\lambda)\widetilde{F}(\lambda)$ is upper triangular with degree k and nonsingular leading coefficient, and the degree of $\widetilde{E}(\lambda)\widetilde{B}(\lambda)$ does not exceed k. Setting

$$E(\lambda) = \begin{bmatrix} \widetilde{E}(\lambda) & 0 \\ 0 & 1 \end{bmatrix}\widehat{E}(\lambda) \quad \text{and} \quad F(\lambda) = \widehat{F}(\lambda)\begin{bmatrix} \widetilde{F}(\lambda) & 0 \\ 0 & 1 \end{bmatrix}$$

then completes the proof. $\qquad\square$

The sparsity preservation property of Möbius transformations described in Remark 9.1, together with the controlled effect of Möbius transformations on Jordan characteristic given by Theorem 5.3, makes it possible to extend the realization result of Lemma 9.2 to situations where both finite *and* infinite elementary divisors may be present. To handle this in a simple fashion, we assume we are given a *list* of elementary divisors rather than a matrix polynomial; indeed, it is convenient to express this list in the form of a specified Jordan characteristic that we wish to realize by some regular matrix polynomial. The following theorem then gives necessary and sufficient conditions for a list of finite and infinite elementary divisors to be realizable by a regular matrix polynomial.

Theorem 9.4 (Regular realization theorem).
Let $\mathbb{F}$ be an algebraically closed field. Furthermore, let $\lambda_1, \ldots, \lambda_m \in \mathbb{F} \cup \{\infty\}$ be pairwise distinct and let

$$(\alpha_{1,i}, \alpha_{2,i}, \ldots, \alpha_{n,i}) \in \mathbb{N}_{\leq}^n$$

for $i = 1, \ldots, m$ be given. Then the following statements are equivalent.

(a) $\displaystyle\sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_{j,i} = kn.$

(b) *There exists a regular matrix polynomial $P(\lambda)$ of size $n \times n$ and grade k such that*

$$\mathcal{J}(P, \lambda_i) = (\alpha_{1,i}, \alpha_{2,i}, \ldots, \alpha_{n,i}) \quad for \quad i = 1, \ldots, m,$$

and $\mathcal{J}(P, \lambda)$ is the zero sequence for all $\lambda \in \mathbb{F}_\infty$ with $\lambda \neq \lambda_i$, $i = 1, \ldots, m$. Moreover, the realizing polynomial $P(\lambda)$ may always be taken to be upper triangular.

Proof. (a $\Rightarrow$ b): Let $A \in GL(2, \mathbb{F})$ be chosen such that the corresponding Möbius function m_A has $\mathsf{m}_A(\lambda_i) \neq \infty$ for all $i = 1, \ldots, m$. (Note that if none of $\lambda_1, \ldots, \lambda_m$ is ∞, then one can simply choose $A = I_2$.) Letting $\mu_i := \mathsf{m}_A(\lambda_i)$, set

$$d_j(\lambda) := \prod_{i=1}^{m} (\lambda - \mu_i)^{\alpha_{j,i}}, \quad j = 1, \ldots, n$$

and $D(\lambda) := \operatorname{diag}\big(d_1(\lambda), \ldots, d_n(\lambda)\big)$. Then $D(\lambda)$ is in Smith form, and by (a) we have $\deg\big(\det D(\lambda)\big) = nk$, so by Lemma 9.2 there exist unimodular matrix polynomials $E(\lambda)$ and $F(\lambda)$ such that $Q(\lambda) := E(\lambda)D(\lambda)F(\lambda)$ is upper triangular, regular, and has degree (and grade) k, with nonsingular leading coefficient. Since $D(\lambda)$ is the Smith form of $Q(\lambda)$, we see that

$$\mathcal{J}\big(Q, \mathsf{m}_A(\lambda_i)\big) = (\alpha_{1,i}, \alpha_{2,i}, \ldots, \alpha_{n,i})$$

for $i = 1, \ldots, m$, and that $\mathcal{J}(Q, \mu)$ is the zero sequence for all $\mu \in \mathbb{F}_\infty$ with $\mu \neq \mathsf{m}_A(\lambda_i)$, $i = 1, \ldots, m$. Then the grade k matrix polynomial $P(\lambda) := \mathbf{M}_A(Q)(\lambda)$ is upper triangular by Remark 9.1, regular by Corollary 3.28, and has the desired Jordan characteristic

$$\mathcal{J}(P, \lambda_i) = \mathcal{J}\big(\mathbf{M}_A(Q), \lambda_i\big) = \mathcal{J}\big(Q, \mathsf{m}_A(\lambda_i)\big) = (\alpha_{1,i}, \alpha_{2,i}, \ldots, \alpha_{n,i})$$

by Theorem 5.3, thus showing that (b) holds.

(b $\Rightarrow$ a): That the sum of the (finite and infinite) partial multiplicities of *any* regular $n \times n$ matrix polynomial is kn seems to be a well-known fact when grade is chosen to be equal to degree. An elementary proof that this sum is kn, valid for any choice of grade k and for regular $n \times n$ matrix polynomials over an arbitrary field, can be found in [23, Lemma 6.1]. $\square$

The result of Theorem 9.4 is also presented in [38] and [61], and extended in each paper to an even more general result. We have included the regular realization theorem and an alternative proof in this paper primarily to highlight the importance of Möbius transformations as a tool for the solution of polynomial inverse eigenproblems.

9.2 Linking Alternating and Palindromic Matrix Polynomials

In [46] and [47] the possible Smith forms of alternating and palindromic matrix polynomials were investigated. Comparing the results of these two papers, an interesting parallelism can be observed. On the one hand, if $D(\lambda) = \operatorname{diag}(d_1(\lambda), \ldots, d_n(\lambda))$ is the Smith form of some T-alternating matrix polynomial, then every $d_i(\lambda)$ is alternating and all elementary divisors

corresponding to zero and infinity satisfy certain pairing conditions [46, Theorem 3.10 and Theorem 3.11]. Similarly, if $D(\lambda)$ is the Smith form of some T-palindromic matrix polynomial, then every $d_i(\lambda)$ is palindromic and all elementary divisors corresponding to $+1$ and -1 satisfy certain pairing conditions [47, Theorem 7.6]. Although slightly different arguments were used to prove these results, the parallelism in the results themselves is not really a surprise, given the previously known fact [44] that T-alternating and T-palindromic matrix polynomials are linked via the Cayley transformation, a special type of Möbius transformation. With the theory developed in this paper, this parallelism is now easy to understand in detail; indeed the main results of [46] can be deduced from the corresponding results in [47], and vice versa, a fact whose simple verification is left to the reader. Instead, we examine the Cayley transformation link between alternating and palindromic matrix polynomials observed in [44], from a slightly different angle. We start by recalling the definitions given in [44] for the convenience of the reader.

Definition 9.5 ([44]). *An $n \times n$ matrix polynomial $P(\lambda)$ of grade k is*

(a) *T-palindromic if $P^T(\lambda) = \operatorname{rev}_k P(\lambda)$,*

(b) *T-anti-palindromic if $P^T(\lambda) = -\operatorname{rev}_k P(\lambda)$,*

(c) *T-even if $P^T(\lambda) = P(-\lambda)$,*

(d) *T-odd if $P^T(\lambda) = -P(-\lambda)$,*

(e) *T-alternating if it is T-even or T-odd.*

As seen in Section 3.1, rev_k is a special case of a Möbius transformation. The same is true for the transformation that maps $P(\lambda)$ to $P(-\lambda)$. Indeed, let

$$ R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}. $$

Then $\mathbf{M}_R(P)(\lambda) = \operatorname{rev}_k P(\lambda)$ and $\mathbf{M}_S(P)(\lambda) = P(-\lambda)$. Therefore the structures in Definition 9.5 can be alternatively expressed in the following manner.

Definition 9.6. *An $n \times n$ matrix polynomial $P(\lambda)$ of grade k is*

(a) *T-palindromic if and only if $P^T = \mathbf{M}_R(P)$,*

(b) *T-anti-palindromic if and only if $P^T = -\mathbf{M}_R(P)$,*

(c) *T-even if and only if $P^T = \mathbf{M}_S(P)$,*

(d) *T-odd if and only if $P^T = -\mathbf{M}_S(P)$.*

Next we recall the theorem from [44] that links T-alternating and T-(anti-)palindromic polynomials via the Cayley transformations $\mathcal{C}_{+1}$ and $\mathcal{C}_{-1}$, introduced in Example 3.10.

Theorem 9.7 ([44], Theorem 2.7). *Let $P(\lambda)$ be a matrix polynomial of grade $k \geq 1$. Then the correspondence between structure in $P(\lambda)$ and in its Cayley transforms is as stated in Table 9.1.*

Keeping in mind that the Cayley transformations $\mathcal{C}_{+1}$ and $\mathcal{C}_{-1}$ correspond to the Möbius transformations $\mathbf{M}_{A_{+1}}$ and $\mathbf{M}_{A_{-1}}$, where A_{+1} and A_{-1} are as in (3.6), we claim that the

Table 9.1: Cayley transformations of structured matrix polynomials

$P(\lambda)$	$\mathcal{C}_{-1}(P)(\mu)$		$\mathcal{C}_{+1}(P)(\mu)$	
	k even	k odd	k even	k odd
T-palindromic	T-even	T-odd	T-even	
T-anti-palindromic	T-odd	T-even	T-odd	
T-even	T-palindromic		T-palindromic	T-anti-palindromic
T-odd	T-anti-palindromic		T-anti-palindromic	T-palindromic

essential content of Table 9.1 reduces to the following multiplicative relationships between the four 2×2 matrices S, R, A_{+1}, and A_{-1}:

$$
\begin{aligned}
A_{+1}\,S &= R\,A_{+1}\,, & (9.2)\\
A_{+1}\,R &= -S\,A_{+1}\,, & (9.3)\\
A_{-1}\,S &= -R\,A_{-1}\,, & (9.4)\\
A_{-1}\,R &= S\,A_{-1}\,. & (9.5)
\end{aligned}
$$

Observe that each of these four equations is just a variation on the statement "R is similar to S" (or to $-S$). Note that the first two equations are related to the $\mathcal{C}_{+1}$ part of Table 9.1, while the last two go with the $\mathcal{C}_{-1}$ part of Table 9.1.

Now let us see in detail how the matrix equation (9.2) "explains" why the Cayley transformation $\mathcal{C}_{+1}$ turns T-palindromic into T-even polynomials. So suppose that $P(\lambda)$ is a T-palindromic polynomial, and

$$
Q(\lambda) := \mathcal{C}_{+1}(P)(\lambda) = \mathbf{M}_{A_{+1}}(P)(\lambda) \tag{9.6}
$$

is the Cayley transform of $P(\lambda)$. Then $P(\lambda)$ being T-palindromic means that $P^T = \mathbf{M}_R(P)$. Inserting this into the transposed equation (9.6), we obtain

$$
\begin{aligned}
Q^T &= \big(\mathbf{M}_{A_{+1}}(P)\big)^T = \mathbf{M}_{A_{+1}}\big(P^T\big) = \mathbf{M}_{A_{+1}}\big(\mathbf{M}_R(P)\big)\\
&= \mathbf{M}_{RA_{+1}}(P) = \mathbf{M}_{A_{+1}S}(P) = \mathbf{M}_S\big(\mathbf{M}_{A_{+1}}(P)\big) = \mathbf{M}_S(Q)\,,
\end{aligned}
$$

where we used Proposition 3.16(a), Theorem 3.18(b), and (9.2). Thus we see that

$$
P^T = \mathbf{M}_R(P) \quad \Longrightarrow \quad Q^T = \mathbf{M}_S(Q)\,, \quad \text{i.e.,} \quad Q^T(\lambda) = Q(-\lambda)\,,
$$

so $Q(\lambda)$ is T-even, as claimed.

As another illustration, consider the last row of Table 9.1, again using the $\mathcal{C}_{+1}$ Cayley transformation. This time let $P(\lambda)$ be T-odd, i.e., $P^T = -\mathbf{M}_S(P)$, and let

$$
Q(\lambda) := \mathcal{C}_{+1}(P) = \mathbf{M}_{A_{+1}}(P)
$$

be its Cayley transform. Then using the properties of Möbius transformations from Propositions 3.16(a) and 3.5, Theorem 3.18(b) and (d), as well as equation (9.3), we obtain that

$$
\begin{aligned}
Q^T &= \big(\mathbf{M}_{A_{+1}}(P)\big)^T = \mathbf{M}_{A_{+1}}\big(P^T\big) = \mathbf{M}_{A_{+1}}\big(-\mathbf{M}_S(P)\big) = -\mathbf{M}_{A_{+1}}\big(\mathbf{M}_S(P)\big)\\
&= -\mathbf{M}_{SA_{+1}}(P) = -\mathbf{M}_{-A_{+1}R}(P) = (-1)^{k+1}\mathbf{M}_R\big(\mathbf{M}_{A_{+1}}(P)\big)\\
&= (-1)^{k+1}\mathbf{M}_R(Q)\,.
\end{aligned}
$$

Thus we easily recover the T-anti-palindromicity of Q when k is even, and the T-palindromicity of Q when k is odd. Analogous calculations produce all the other entries in Table 9.1.

In summary, then, the perspective afforded by our results about general Möbius transformations enables us to transparently see the Cayley connection between alternating and palindromic structure as a straightforward consequence of simple relations between 2×2 matrices.

10   Conclusions

The wide range of properties established in this paper for Möbius transformations acting on general matrix polynomials over an arbitrary field $\mathbb{F}$ make them a potent tool for both discovering and proving nontrivial results.

Since a Möbius transformation can alter the degree of a polynomial, we have used polynomials of fixed grade as the domain of discourse. We can then view Möbius transformations as a group of $\mathbb{F}$-linear operators on a finite-dimensional vector space, and consequently develop a theory that is simpler and more unified than it would otherwise be. Many important transformations are now seen to be special instances of Möbius transformations.

We have shown that a number of structural features of matrix polynomials are preserved by Möbius transformations, for example, regularity, rank, minimal indices, the location of zero entries, as well as symmetry and skew-symmetry. The relations of strict equivalence and spectral equivalence and the property of being a strong linearization or strong quadratification are also preserved. Other important features like the determinant, Jordan characteristic, Smith form, invariant pairs and minimal bases have each been shown to change in a simple and predictable way.

By linking palindromic and alternating matrix polynomials via Möbius transformations, the parallelism between the pairing conditions on their elementary divisors can be understood and even predicted. Möbius transformations are thus shown to be an efficient device for transferring results between matrix polynomial structures.

Finally, their role in yielding necessary and sufficient conditions for a given list of finite and infinite elementary divisors to be realizable by a regular matrix polynomial highlights the potential of Möbius transformations as effective tools for solving inverse polynomial eigenvalue problems.

Acknowledgments

References

[1] L.V. Ahlfors. *Complex Analysis*. McGraw-Hill, New York, 3rd edition, 1979.

[2] M. Al-Ammari and F. Tisseur. Hermitian matrix polynomials with real eigenvalues of definite type. Part I: Classification. *Linear Algebra Appl.*, 436:3954–3973, 2012.

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

[4] A. Amparan, S. Marcaida, and I. Zaballa. An interpretation of Rosenbrock's theorem via local rings. In Hassan Abid Yasser, editor, *Linear Algebra - Theorems and Applications*. InTech, 2012. DOI: 10.5772/46483. Available from:

http://www.intechopen.com/books/linear-algebra-theorems-and-applications/an-interpretation-of-rosenbrock-s-theorem-via-local-rings.

[5] A. Amparan, S. Marcaida, and I. Zaballa. On the structure invariants of proper rational matrices with prescribed finite poles. *Linear Multilin. Algebra*, 61:1464–1486, 2013.

[6] A. Amparan, S. Marcaida, and I. Zaballa. Finite and infinite structures of rational matrices: a local approach. Submitted for publication, 2014.

[7] K.J. Astrom. *Computer Controlled Systems, Theory and Design*. Prentice-Hall, Englewood Cliffs, New Jersey, second edition, 1999.

[8] J. Ball, I. Gohberg, and L. Rodman. *Interpolation of rational matrix functions*, volume 45 of *Operator Theory: Advances and Applications*. Birkhäuser, Basel, Switzerland, 1990.

[9] T. Betcke and D. Kressner. Perturbation, extraction and refinement of invariant pairs for matrix polynomials. *Linear Algebra Appl.*, 435:514–536, 2011.

[10] W.-J. Beyn and V. Thümmler. Continuation of invariant subspaces for parameterized quadratic eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 31:1361–1381, 2009.

[11] D. Bini, B. Iannazzo, and B. Meini. *Numerical Solution of Algebraic Riccati Equations*. Fundamentals of Algorithms 09. SIAM, Philadelphia, PA, 2012.

[12] G. Birkhoff and S. MacLane. *A Survey of Modern Algebra*. MacMillan, New York, 2nd edition, 1953.

[13] R. Byers. *Hamiltonian and Symplectic Algorithms for the Algebraic Riccati Equation*. PhD thesis, Cornell University, Dept. Comp. Sci., Ithaca, NY, 1983.

[14] R. Byers, V. Mehrmann, and H. Xu. Trimmed linearizations for structured matrix polynomials. *Linear Algebra Appl.*, 429:2373–2400, 2008.

[15] I. Cabral and F. C. Silva. Unified theorems on completions of matrix pencils. *Linear Algebra Appl.*, 159:43–54, 1991.

[16] E. Chitambar, C. Miller, and Y. Shi. Matrix pencils and entanglement classification. *J. Math. Physics*, 51:072205, 2010.

[17] J. B. Conway. *Functions of one complex variable*. Springer-Verlag, New York, 2nd edition, 1978.

[18] R. M. Corless, N. Rezvani, and A. Amiraslani. Pseudospectra of matrix polynomials that are expressed in alternative bases. *Math. in Computer Sci.*, 1:353–374, 2007.

[19] F. De Terán and F. Dopico. Sharp lower bounds for the dimension of linearizations of matrix polynomials. *Electron. J. Linear Algebra*, 17:518–531, 2008.

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

[21] F. De Terán, F. Dopico, and D. S. Mackey. Fiedler companion linearizations and the recovery of minimal indices. *SIAM J. Matrix Anal. Appl.*, 31:2181–2204, 2010.

[22] F. De Terán, F. Dopico, and D. S. Mackey. Fiedler companion linearizations for rectangular matrix polynomials. *Linear Algebra Appl.*, 437:957–991, 2012.

[23] F. De Terán, F. Dopico, and D. S. Mackey. Spectral equivalence of matrix polynomials and the index sum theorem. To appear in *Linear Algebra Appl.* Also available as MIMS EPrint 2013.47, Manchester Institute for Mathematical Sciences, The University of Manchester, UK., 2013.

[24] J. Dion and C. Commault. Smith-McMillan factorizations at infinity of rational matrix functions and their control interpretation. *Sys. Control Lett.*, 1:312–320, 1982.

[25] G. D. Forney. Minimal bases of rational vector spaces, with applications to multivariable linear systems. *SIAM J. Cont.*, 13:493–520, 1975.

[26] G. Frobenius. Theorie der linearen Formen mit ganzen Coefficienten. *J. Reine Angew. Math. (Crelle)*, 86:146–208, 1878.

[27] F. R. Gantmacher. *Theory of Matrices.* Chelsea, New York, 1959.

[28] J. D. Gardiner and A. J. Laub. A generalization of the matrix-sign-function solution for algebraic Riccati equations. *Internat. J. Control*, 44:823–832, 1986.

[29] I. Gohberg, M. A. Kaashoek, and P. Lancaster. General theory of regular matrix polynomials and band Toeplitz operators. *Integral Equations Operator Theory*, 11:776–882, 1988.

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

[31] C.-H. Guo, B. Iannazzo, and B. Meini. On the doubling algorithm for a (shifted) non-symmetric algebraic Riccati equation. *SIAM J. Matrix Anal. Appl.*, 29:1083–1100, 2008.

[32] C.-H. Guo and W.-W. Lin. Solving a structured quadratic eigenvalue problem by a structure-preserving doubling algorithm. *SIAM J. Matrix Anal. Appl.*, 31:2784–2801, 2010.

[33] W. Hahn. *Stability of Motion.* Springer-Verlag, Heidelberg, 1967.

[34] E. Hairer, C. Lubich, and G. Wanner. *Numerical Geometric Integration.* Springer-Verlag, Heidelberg, 2002.

[35] N. J. Higham, D. S. Mackey, and F. Tisseur. Definite matrix polynomials and their linearizations by definite pencils. *SIAM J. Matrix Anal. Appl.*, 31:478–502, 2009.

[36] D. Hinrichsen and A. J. Pritchard. *Mathematical Systems Theory I: Modelling, State Space Analysis, Stability and Robustness.* Springer-Verlag, Heidelberg, 2005.

[37] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1985.

[38] S. Johansson, B. Kågström, and P. Van Dooren. Stratification of full rank polynomial matrices. *Linear Algebra Appl.*, 439:1062–1090, 2013.

[39] B. Kågström and A. Ruhe, editors. *Matrix Pencils*, volume 973 of *Lecture Notes in Math.* Springer-Verlag, Berlin, 1983.

[40] P. Lancaster. Linearization of regular matrix polynomials. *Electron. J. Linear Algebra*, 17:21–27, 2008.

[41] P. Lancaster and P. Psarrakos. A note on weak and strong linearizations of regular matrix polynomials. MIMS EPrint 2006.72, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2005.

[42] P. Lancaster and L. Rodman. *The Algebraic Riccati Equation*. Oxford University Press, Oxford, 1995.

[43] D.S. Mackey. Minimal indices and minimal bases via filtrations. MIMS EPrint 2012.82, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2012. Submitted for publication.

[44] D.S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Structured polynomial eigenvalue problems: Good vibrations from good linearizations. *SIAM J. Matrix Anal. Appl.*, 28:1029–1051, 2006.

[45] D.S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Vector spaces of linearizations for matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 28:971–1004, 2006.

[46] D.S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Jordan structures of alternating matrix polynomials. *Linear Algebra Appl.*, 432:867–891, 2010.

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

[48] D.S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Skew-symmetric matrix polynomials and their Smith forms. *Linear Algebra Appl.*, 438:4625–4653, 2013.

[49] M. Marcus and H. Minc. *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, Boston, 1964.

[50] V. Mehrmann. *The Autonomous Linear Quadratic Control Problem, Theory and Numerical Solution*. Number 163 in Lecture Notes in Control and Information Sciences. Springer-Verlag, Heidelberg, July 1991.

[51] V. Mehrmann. A step toward a unified treatment of continuous and discrete time control problems. *Linear Algebra Appl.*, 241–243:749–779, 1996.

[52] V. Mehrmann and F. Poloni. A generalized structured doubling algorithm for the numerical solution of linear quadratic optimal control problems. *Numer. Lin. Alg. Appl.*, 20:112–137, 2013.

[53] V. Noferini. The behaviour of the complete eigenstructure of a polynomial matrix under a generic rational transformation. *Electron. J. Linear Algebra*, 23:607–624, 2012.

[54] R. Ober and S. Montgomery-Smith. Bilinear transformation of infinite dimensional state space systems and balanced realizations of nonrational transfer functions. *SIAM J. Cont. Optim.*, 28:438–465, 1990.

[55] A. Oppenheim. *Discrete Time Signal Processing*. Pearson Higher Education, Upper Saddle River, New Jersey, third edition, 2010.

[56] G. Polya and G. Latta. *Complex Variables*. Wiley, New York, 1974.

[57] V. V. Prasolov. *Problems and Theorems in Linear Algebra*. American Mathematical Society, Providence, 1994.

[58] R. Remmert. *Theory of Complex Functions*, volume 122 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991.

[59] V. Sima. *Algorithms for Linear-Quadratic Optimization*, volume 200 of *Pure and Applied Mathematics*. Marcel Dekker, Inc., New York, NY, 1996.

[60] G. W. Stewart. On the sensitivity of the eigenvalue problem $Ax = \lambda Bx$. *SIAM J. Numer. Anal.*, 9(4):669–686, 1972.

[61] L. Taslaman, F. Tisseur, and I. Zaballa. Triangularizing matrix polynomials. *Linear Algebra Appl.*, 439:1679–1699, 2013.

[62] F. Tisseur and I. Zaballa. Triangularizing quadratic matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 34:312–337, 2013.

[63] K. Vahlen. Über Bewegungen und complexe Zahlen. *Math. Ann*, 55:585–593, 1902.

[64] A. I. G. Vardulakis. *Linear Multivariable Control*. Wiley, New York, 1991.

[65] H. K. Wimmer. The structure of nonsingular polynomial matrices. *Math. Systems Theory*, 14:367–379, 1981.

[66] I. Zaballa and F. Tisseur. Finite and infinite elementary divisors of matrix polynomials: a global approach. MIMS EPrint 2012.78, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2012.