MANCHESTER 1824

The University of Manchester

Structured Linearizations for Matrix Polynomials

Mackey, D. Steven

2006

MIMS EPrint: **2006.68**

ISSN 1749-9097

STRUCTURED LINEARIZATIONS FOR MATRIX POLYNOMIALS

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2006

D. Steven Mackey
School of Mathematics

Contents

List of Tables

List of Figures

Abstract

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

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

Many applications give rise to nonlinear eigenvalue problems with an underlying structured matrix polynomial; perhaps the most well-known are symmetric and Hermitian polynomials. In this thesis we also identify several less well-known types of structured polynomial (e.g., palindromic, even, odd), explore the relationships between them, and illustrate their appearance in a variety of applications. Special classes of linearizations that reflect the structure of these polynomials, and therefore preserve symmetries in their spectra, are introduced and investigated. We analyze the existence and uniqueness of such linearizations, and show how they may be systematically constructed.

The infinitely many linearizations of any given polynomial P can have widely varying eigenvalue condition numbers. We investigate the conditioning of linearizations from $\mathbb{DL}(P)$, looking for the best conditioned linearization in that space and comparing its conditioning with that of the original polynomial. We also analyze the eigenvalue conditioning of the widely used first and second companion linearizations, and find that they can potentially be much more ill conditioned than P. Our results are phrased in terms of both the standard relative condition number and the condition number of Dedieu and Tisseur for the problem in homogeneous form, this latter condition number having the advantage of applying to zero and infinite eigenvalues.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Copyright

Publications

The work in this thesis is based on the contents of four publications.

Chapters 2 and 4 are based on the paper "Vector Spaces of Linearizations for Matrix Polynomials" [58] (with N. Mackey, C. Mehl, and V. Mehrmann), to appear in SIAM Journal of Matrix Analysis and Applications.

Chapters 3 and 5 are based on the technical report "Symmetric Linearizations for Matrix Polynomials" [34] (with N.J. Higham, N. Mackey, and F. Tisseur), MIMS EPrint 2005.25, Manchester Institute for Mathematical Sciences, Jan. 2006. This work has been submitted for publication in SIAM Journal of Matrix Analysis and Applications.

Chapters 6 and 7 are based on the technical report "Structured Polynomial Eigenvalue Problems: Good Vibrations from Good Linearizations" [57] (with N. Mackey, C. Mehl, and V. Mehrmann), MIMS EPrint 2006.38, Manchester Institute for Mathematical Sciences, Mar. 2006, and its predecessor "Palindromic Polynomial Eigenvalue Problems", Apr. 2005. This work has been submitted for publication in SIAM Journal of Matrix Analysis and Applications.

Finally, chapter 8 is based on the paper "The Conditioning of Linearizations of Matrix Polynomials" [35] (with N.J. Higham and F. Tisseur), to appear in SIAM Journal of Matrix Analysis and Applications.

Acknowledgements

I'd like to thank all the family, friends, and colleagues who by their support and encouragement have helped me reach this point after so many years.

My first debt of gratitude goes to my parents, Don and Eula Mackey, who gave me the freedom to pursue my own thoughts and interests as I was growing up, even when they led to places where we didn't see eye to eye. For their enduring love and support, as well as that from my parents-in-law Kamela and Vishram Masojee in faraway Bombay, I thank them.

I am also grateful for having the chance to do mathematics with Nil Mackey in Kalamazoo, Christian Mehl and Volker Mehrmann in Berlin, and Nick Higham and Fran Tisseur in Manchester, who each played a key role in the development that led to this thesis. It is hard to imagine a finer group of people to be able to put one's name next to. Special thanks, though, are due to Nick and Fran for everything they have done to make it possible for this work to exist in the form of a thesis.

Beyond all gratitude and thanks is the debt I owe to Nil — colleague, friend, drill sergeant, and helpmate. For all the patience over the years and enduring belief in my ability to do something of value, I dedicate this thesis to her.

Chapter 1

Introduction

Polynomial eigenvalue problems $P(\lambda)x = 0$, where $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ with real or complex coefficient matrices A_i, form the basis for (among many other applications) the vibration analysis of buildings, machines, and vehicles [31], [47], [83], and numerical methods for the solution of these problems are incorporated into most commercial and non-commercial software packages for structural analysis.

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

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

Yet many linearizations exist, and other than the convenience of their construction, there is no apparent reason for preferring the companion forms. Indeed one obvious disadvantage is their lack of preservation of structural properties of P like symmetry. But physical problems often lead to matrix polynomials that are structured in some way; for example, the coefficient matrices may all be symmetric [47], or perhaps alternate between symmetric and skew-symmetric [65], or even have palindromic structure [57]. Such structure in a matrix polynomial often forces symmetries or constraints on its spectrum [57], [64], [65], [83] that have physical significance. Numerical methods (in a finite precision environment) that ignore this structure can destroy these qualitatively important spectral symmetries, sometimes even to the point of producing physically meaningless or uninterpretable results [83]. Since the companion form linearizations do not reflect any structure that may be present in the original polynomial, their use for numerical computation in such situations may be problematic. Therefore it is important to be able to construct linearizations that reflect the structure of the given matrix polynomial, and then develop numerical methods for the corresponding linear eigenvalue problem that properly address these structures as well. The latter topic has been an important area of research in the last decade, see, e.g., [8], [16], [24], [61], [64], [65] and the references therein.

An important issue for any computational problem is its conditioning, i.e., its sensitivity to small perturbations. It is known that different linearizations for a given polynomial eigenvalue problem can have very different conditioning [80], [83], so that numerical methods may produce rather different results for each linearization. It would clearly be useful to have available a large class of easily constructible linearizations from which one could always select a linearization guaranteed to be as well-conditioned as the original problem.

A further issue for linearizations concerns eigenvalues at ∞. Much of the literature on polynomial eigenvalue problems considers only polynomials whose leading coefficient matrix A_k is nonsingular (or even the identity), so the issue of infinite eigenvalues doesn't even arise. But there are a number of applications, such as constraint multi-body systems [22], [73], circuit simulation [27], or optical waveguide design [74], where the leading coefficient is singular. In such cases one must choose a linearization with care, since not all linearizations properly reflect the Jordan structure of the eigenvalue ∞ [63]. It has therefore been suggested [30], [49] that only *strong linearizations*, which are guaranteed to preserve the structure of infinite eigenvalues, can safely be used in such circumstances. Having a large class of linearizations that are known to also be strong linearizations would make this issue of infinite eigenvalues less of a concern in practice.

The first major aim of this thesis is to show how to systematically generate two large classes of linearizations that address these issues, thereby broadening the menu of linearizations that are readily available for computations. The linearizations in these classes are easy to construct from the data in P, properly handle any infinite eigenvalues, provide a fertile source of structured linearizations for many types of structured polynomials [34], [57], and collectively constitute a well-defined arena in which to look for "optimally" conditioned linearizations [35].

Taking the two companion forms as prototypes, we begin in Chapter 2 by showing how to associate to a *general* matrix polynomial P two large vector space of pencils, denoted by $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. The basic properties of these pencils are then developed, and almost all of them are found to be linearizations for P. Special subspaces of $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ are then explored in even greater detail in Chapters 3 and 4. These pencil spaces are the arenas in which the rest of the thesis is played out.

The second major theme of this work is structured matrix polynomials, and the preservation of this structure in linearizations. The two main types of structure considered here, indeed the ones that originally motivated the investigation leading to this thesis, are symmetric and palindromic structure. We show in Chapters 5–7 that the pencil spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ developed in Chapters 2–4 for general polynomials are rich enough to include subspaces of pencils that reflect symmetric or palindromic structure, whenever the polynomial P has one of those structures.

Symmetric polynomials are of course well recognized as being important for applications, particularly for ones involving the vibration analysis of mechanical systems [47]. But why is it of any interest to preserve symmetry when we linearize such a polynomial? A matrix polynomial that is real symmetric or Hermitian has a spectrum that is symmetric with respect to the real axis, and the sets of left and right eigenvectors coincide. These properties are preserved in any symmetric (Hermitian) linearization *by virtue of its structure*—not just through the numerical entries of the pencil. A symmetry-preserving pencil also has the practical advantages that storage and computational costs are reduced if a method that exploits symmetry is applied. The eigenvalues of a symmetric (Hermitian) pencil $L(\lambda) = \lambda X + Y$ can be computed, for small to medium size problems, by first reducing the matrix pair (Y, X) to tridiagonal-diagonal form [81] and then using the HR [11], [14] or LR [72] algorithms or the Ehrlich-Aberth iterations [10]. For large problems, a symmetry-preserving pseudo-Lanczos algorithm of Parlett and Chen [68], [6, Sec. 8.6], based on an indefinite inner product, can be used. For a quadratic polynomial $Q(\lambda)$ that is hyperbolic, or in particular overdamped, a linearization that is a symmetric definite pencil can be identified [36, Thm. 3.6]; this pencil is amenable to structure-preserving methods that exploit both the symmetry and the definiteness [85], and guarantee real computed eigenvalues for any $Q(\lambda)$ that is not too close to being non-hyperbolic.

Palindromic matrix polynomials, on the other hand, arise in a variety of applications, but heretofore have not been widely appreciated as a significant class of problems worthy of separate analysis. Our main motivation for isolating this structure and studying it in some detail comes from a project investigating the rail traffic noise caused by high speed trains [37], [38]. The eigenvalue problem that arises in this project has the form

$$(\lambda^2 A + \lambda B + A^T)x = 0\,, \tag{1.0.1}$$

where A, B are complex square matrices with B complex symmetric and A singular. Observe that the matrix polynomial in (1.0.1) has the property that reversing the order of the coefficient matrices, followed by taking their transpose, leads back to the original matrix polynomial. By analogy with linguistic palindromes, of which

Sex at noon taxes

is perhaps a less well-known example[1], we call such matrix polynomials *T-palindromic.*

Quadratic real and complex T-palindromic eigenvalue problems also arise in the mathematical modelling and numerical simulation of the behavior of periodic surface acoustic wave (SAW) filters [89], whereas the computation of the Crawford number [36] associated with the perturbation analysis of symmetric generalized eigenvalue problems produces a quadratic $*$-palindromic eigenvalue problem, where $*$ denotes conjugate transpose. Higher order matrix polynomials with $*$-palindromic structure also arise in problems of discrete optimal control [57].

Alternating matrix polynomials, i.e. polynomials whose matrix coefficients alternate between symmetric and skew-symmetric, forms another significant class of structured polynomial that deserves wider recognition. Although seemingly unrelated, palindromic and alternating structures turn out to be intimately connected with each other via a matrix polynomial version of the Cayley transformation. This Cayley connection is just one aspect of an interesting analogy that develops in Chapter 6 between palindromic and alternating polynomials on the one hand, and symplectic and Hamiltonian matrices on the other. Indeed, we find in Chapters 6 and 7 that alternating polynomials can be studied side-by-side with palindromic polynomials, and their properties and structured linearizations developed using similar techniques.

[1] Invented by the mathematician Peter Hilton in 1947 for his advisor J.H.C. Whitehead. (It is probable, Hilton says, that this palindrome was known before 1947.) When Whitehead lamented its brevity, Hilton responded [39] by crafting the palindromic masterpiece "Doc, note, I dissent. A fast never prevents a fatness. I diet on cod." [53, p. 287] A much longer palindrome has recently been discovered—a section of the DNA sequence in the human male Y chromosome. [71] [77] [87]

More details concerning applications of alternating and palindromic polynomials can be found in Section 6.4.

The profusion of pencils provided in Chapters 2–7 that potentially linearize a polynomial P poses a possible problem for any potential user: how to choose which linearization to use? The third (and final) goal of this thesis is to provide some guidance on answering this question via an analysis of the conditioning of the eigenvalues of linearizations in Chapter 8. The focus in this analysis is on pencils in the intersection $\mathbb{L}_1(P) \cap \mathbb{L}_2(P) =: \mathbb{DL}(P)$; the fact that both the left and right eigenvectors of any pencil in $\mathbb{DL}(P)$ are simply related to the left and right eigenvectors of P makes it possible to directly compare the conditioning of any $\mathbb{DL}(P)$-linearization with the conditioning of P. Moreover, the special properties of $\mathbb{DL}(P)$-pencils allow one to identify a "near optimally" conditioned pencil in $\mathbb{DL}(P)$ for any given eigenvalue of P. Because of the central role of the companion forms in current computational practice, we also give a separate analysis of the conditioning of these particular linearizations. The resulting formulas reveal the possibility of instability of the companion forms in certain circumstances. Some numerical experiments illustrating the efficacy of our analysis are presented in Section 8.7 to conclude the thesis.

1.1 Some Preliminaries

In this section we establish some of the basic definitions and notational conventions to be used throughout this thesis. The fundamental objects of study are $n \times n$ matrix polynomials of the form

$$P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i \,, \quad A_0, \,\dots\,, A_k \in \mathbb{F}^{n\times n}, \; A_k \neq 0 \,, \tag{1.1.1}$$

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

Definition 1.1.1. If $\lambda \in \mathbb{C}$ and nonzero $x \in \mathbb{C}^n$ satisfy $P(\lambda)x = 0$, then x is said to be a *right eigenvector* of P corresponding to the (finite) eigenvalue λ.

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

Our main concern is with *regular matrix polynomials*, i.e., polynomials $P(\lambda)$ such that $\det P(\lambda)$ is not identically zero for all $\lambda \in \mathbb{C}$; for such polynomials the finite eigenvalues are precisely the roots of the scalar polynomial $\det P(\lambda)$. Note, however, that some of our results also hold for singular matrix polynomials (these are studied in detail in [63], [76]).

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

Definition 1.1.2 (Reversal of matrix polynomials). For a matrix polynomial $P(\lambda)$ of degree k as in (1.1.1), the *reversal* of $P(\lambda)$ is the polynomial

$$\operatorname{rev} P(\lambda) := \lambda^k P(1/\lambda) = \sum_{i=0}^{k} \lambda^i A_{k-i} \,. \tag{1.1.2}$$

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

Definition 1.1.3 (Eigenvalue at ∞). Let $P(\lambda)$ be a regular matrix polynomial of degree $k \geq 1$. Then $P(\lambda)$ is said to have an eigenvalue at ∞ with eigenvector x if $\operatorname{rev} P(\lambda)$ has the eigenvalue 0 with eigenvector x. The algebraic, geometric, and partial multiplicities of the infinite eigenvalue are defined to be the same as the corresponding multiplicities of the zero eigenvalue of $\operatorname{rev} P(\lambda)$.

The classical approach to solving and investigating polynomial eigenvalue problems $P(\lambda)x = 0$ is to first perform a *linearization*, that is, to transform the given polynomial into a linear matrix pencil $L(\lambda) = \lambda X + Y$ with the same eigenvalues, and then work with this pencil. This transformation of polynomials to pencils is mediated by *unimodular* matrix polynomials[2], i.e., matrix polynomials $E(\lambda)$ such that $\det E(\lambda)$ is a nonzero constant, independent of λ.

Definition 1.1.4 (Linearization [31]). Let $P(\lambda)$ be an $n \times n$ matrix polynomial of degree k with $k \geq 1$. A pencil $L(\lambda) = \lambda X + Y$ with $X, Y \in \mathbb{F}^{kn \times kn}$ is called a *linearization* of $P(\lambda)$ if there exist unimodular matrix polynomials $E(\lambda), F(\lambda)$ such that

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

Note that an immediate consequence of this definition is that $\gamma \det(L(\lambda)) = \det(P(\lambda))$ for some nonzero constant γ, so that L and P have the same spectrum.

There are many different possibilities for linearizations, but probably the most important examples in practice have been the so-called companion forms [51, Sec. 14.1] or companion polynomials [31].

Definition 1.1.5 (Companion Forms). If we let

$$X_1 = X_2 = \operatorname{diag}(A_k, I_{(k-1)n}), \tag{1.1.3a}$$

$$Y_1 = \begin{bmatrix} A_{k-1} & A_{k-2} & \cdots & A_0 \\ -I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -I_n & 0 \end{bmatrix}, \quad \text{and} \quad Y_2 = \begin{bmatrix} A_{k-1} & -I_n & \cdots & 0 \\ A_{k-2} & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & -I_n \\ A_0 & 0 & \cdots & 0 \end{bmatrix}, \tag{1.1.3b}$$

then $C_1(\lambda) = \lambda X_1 + Y_1$ and $C_2(\lambda) = \lambda X_2 + Y_2$ are respectively called the *first* and *second companion forms* for $P(\lambda)$ in (1.1.1).

The notion of linearization in Definition 1.1.4 has been designed mainly for matrix polynomials (1.1.1) with invertible leading coefficient A_k. In this case all the eigenvalues of $P(\lambda)$ are finite, and their Jordan structures (i.e., their partial multiplicities) may be recovered from *any* linearization [31]. However, the situation is somewhat different when the leading coefficient of a regular $P(\lambda)$ is singular, so that ∞ is an

[2]Unimodular matrix polynomials can also be characterized as the "invertible" matrix polynomials, i.e., those with an inverse that is also a matrix polynomial.

eigenvalue with some multiplicity $m > 0$. Although the Jordan structures of all the finite eigenvalues of P are still faithfully recovered from any linearization of P, the eigenvalue ∞ is problematic. Consider, for example, the fact that the identity matrix is a linearization for any unimodular $P(\lambda)$. Indeed, in [49] it is shown that *any* Jordan structure for the eigenvalue ∞ that is compatible with its algebraic multiplicity m can be realized by some linearization for P. Thus linearization in the sense of Definition 1.1.4 completely fails to reflect the Jordan structure of infinite eigenvalues.

To overcome this deficiency, a modification of Definition 1.1.4 was introduced in [30], and termed *strong linearization* in [49]. The correspondence between the infinite eigenvalue of a matrix polynomial P and the eigenvalue zero of $\mathrm{rev}\, P$ is the source of this strengthened definition.

Definition 1.1.6 (Strong Linearization). Let $P(\lambda)$ be a matrix polynomial of degree k with $k \geq 1$. If $L(\lambda)$ is a linearization for $P(\lambda)$ and $\mathrm{rev}\, L(\lambda)$ is a linearization for $\mathrm{rev}\, P(\lambda)$, then $L(\lambda)$ is said to be a *strong linearization* for $P(\lambda)$.

For regular polynomials $P(\lambda)$, the additional property that $\mathrm{rev}\, L(\lambda)$ is a linearization for $\mathrm{rev}\, P(\lambda)$ ensures that the Jordan structure of the eigenvalue ∞ is preserved by strong linearizations. The first and second companion forms of any regular polynomial P have this additional property [30], and thus are always strong linearizations for P. Most of the pencils we construct in this thesis will be shown to be strong linearizations.

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

$$R = R_k = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}, \quad \text{and} \quad N = N_k = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}. \tag{1.1.4}$$

The vector $\begin{bmatrix} \lambda^{k-1} & \lambda^{k-2} & \cdots & \lambda & 1 \end{bmatrix}^T \in \mathbb{F}^k$ of decreasing powers of λ is denoted by Λ. We will also sometimes use Λ with an argument, so that

$$\Lambda(r) := \begin{bmatrix} r^{k-1} & r^{k-2} & \cdots & r & 1 \end{bmatrix}^T. \tag{1.1.5}$$

Denoting the Kronecker product by $\otimes$, the unimodular matrix polynomials

$$T(\lambda) = \begin{bmatrix} 1 & \lambda & \lambda^2 & \cdots & \lambda^{k-1} \\ & 1 & \lambda & \ddots & \vdots \\ & & 1 & \ddots & \lambda^2 \\ & & & \ddots & \lambda \\ & & & & 1 \end{bmatrix} \otimes I \quad \text{and} \quad G(\lambda) = \begin{bmatrix} 1 & & & \lambda^{k-1} \\ & \ddots & & \vdots \\ & & 1 & \lambda \\ & & & 1 \end{bmatrix} \otimes I \tag{1.1.6}$$

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

$$T(\lambda) = G(\lambda) \begin{bmatrix} I & \lambda I & & & \\ & I & & & \\ & & I & & \\ & & & \ddots & \\ & & & & I \end{bmatrix} \begin{bmatrix} I & & & & \\ & I & \lambda I & & \\ & & I & & \\ & & & \ddots & \\ & & & & I \end{bmatrix} \cdots \begin{bmatrix} I & & & & \\ & \ddots & & & \\ & & I & \lambda I & \\ & & & I & \\ & & & & I \end{bmatrix}. \tag{1.1.7}$$

Chapter 2

A Vector Space Setting for Structured Linearizations

The first goal of this thesis is to show how to systematically associate to any given matrix polynomial P two large vector spaces of easily constructible pencils, denoted $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. We will see that these pencils, simply by virtue of being elements of these spaces, are all very close to being linearizations for P. Thus these spaces provide a "sehr bequeme" setting[1] in which to search for linearizations with "extra" properties, in particular structure preservation. In later chapters it will indeed be seen that these spaces are fertile sources of structured linearizations for many kinds of structured matrix polynomial.

2.1 Vector Spaces of "Potential" Linearizations

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

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

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

To motivate the definition of these spaces, let us recall the origin of the first companion form. Imitating the standard procedure for converting a system of higher order linear differential algebraic equations into a first order system (see [31]), introduce the vector variables

$$x_1 = \lambda^{k-1}x,\ x_2 = \lambda^{k-2}x,\ \dots,\ x_{k-1} = \lambda x,\ x_k = x\,, \tag{2.1.1}$$

[1] Apologies to Jacobi.

into the $n \times n$ polynomial eigenvalue problem $P(\lambda)x = \left(\sum_{i=0}^{k} \lambda^i A_i\right)x = 0$, thereby transforming it into

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

Then, together with the relations (2.1.1) between successive variables, this can all be expressed as the $kn \times kn$ linear eigenvalue problem

$$\underbrace{\left(\lambda \begin{bmatrix} A_k & 0 & \cdots & 0 \\ 0 & I_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_n \end{bmatrix} + \begin{bmatrix} A_{k-1} & A_{k-2} & \cdots & A_0 \\ -I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -I_n & 0 \end{bmatrix} \right)}_{=\, C_1(\lambda)} \begin{bmatrix} x_1 \\ \vdots \\ x_{k-1} \\ x_k \end{bmatrix} = 0\,. \tag{2.1.2}$$

Conversely, if we start with (2.1.2), then the last $k-1$ block rows immediately constrain any solution of (2.1.2) to have the form

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

for some vector $x \in \mathbb{F}^n$. Thus to solve (2.1.2) it is reasonable to restrict attention to products of the form $C_1(\lambda) \cdot \big(\Lambda \otimes x\big)$. But

$$C_1(\lambda) \cdot \big(\Lambda \otimes x\big) = \begin{bmatrix} \big(P(\lambda)x\big)^T & 0 & \cdots & 0 \end{bmatrix}^T \quad \text{for } \textit{all } \; x \in \mathbb{F}^n, \tag{2.1.3}$$

and so any solution of (2.1.2) leads to a solution of the original problem $P(\lambda)x = 0$. Now observe that (2.1.3) with its "for all x" quantifier is equivalent to the single identity

$$C_1(\lambda) \cdot \big(\Lambda \otimes I_n\big) = C_1(\lambda) \begin{bmatrix} \lambda^{k-1} I_n \\ \vdots \\ \lambda I_n \\ I_n \end{bmatrix} = \begin{bmatrix} P(\lambda) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = e_1 \otimes P(\lambda)\,. \tag{2.1.4}$$

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

$$L(\lambda) \cdot (\Lambda \otimes I_n) = L(\lambda) \begin{bmatrix} \lambda^{k-1} I_n \\ \vdots \\ \lambda I_n \\ I_n \end{bmatrix} = \begin{bmatrix} v_1 P(\lambda) \\ \vdots \\ v_{k-1} P(\lambda) \\ v_k P(\lambda) \end{bmatrix} = v \otimes P(\lambda) \tag{2.1.5}$$

for some vector $v = [v_1, \cdots, v_k]^T \in \mathbb{F}^k$. This set of pencils will be denoted by $\mathbb{L}_1(P)$ as a reminder that it generalizes the *first* companion form of P. To work with property (2.1.5) more effectively we also introduce the notation

$$\mathcal{V}_P = \{v \otimes P(\lambda) \,:\, v \in \mathbb{F}^k\} \tag{2.1.6}$$

for the set of all possible right-hand sides of (2.1.5). Thus we have the following definition.

Definition 2.1.1. $\mathbb{L}_1(P) := \Big\{ L(\lambda) = \lambda X + Y \,:\, X, Y \in \mathbb{F}^{kn\times kn},\ L(\lambda)\cdot(\Lambda \otimes I_n) \in \mathcal{V}_P \Big\}$.

We will sometimes use the phrase "$L(\lambda)$ satisfies the right ansatz with vector v" or "v is the right ansatz vector for $L(\lambda)$" when $L(\lambda) \in \mathbb{L}_1(P)$ and the vector v in (2.1.5) is the focus of attention. We say "right" ansatz here because $L(\lambda)$ is multiplied on the right by the block column $\Lambda \otimes I_n$; later we introduce an analogous "left ansatz".

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

Proposition 2.1.2. *For any polynomial $P(\lambda)$, $\mathbb{L}_1(P)$ is a vector space over $\mathbb{F}$.*

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

Our next goal is to show that, like the companion forms, pencils in $\mathbb{L}_1(P)$ are easily constructible from the data in P. A consequence of this construction is a characterization of all the pencils in $\mathbb{L}_1(P)$, and a calculation of $\dim \mathbb{L}_1(P)$. To simplify the discussion, we introduce the following new operation on block matrices as a convenient tool for working with products of the form $L(\lambda)\cdot(\Lambda \otimes I_n)$.

Definition 2.1.3 (Column Shifted Sum). Let X and Y be block $k \times k$ matrices

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

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

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

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

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

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

Thus, the property $C_1(\lambda)\cdot(\Lambda \otimes I_n) = e_1 \otimes P(\lambda)$ from (2.1.4) translates in terms of the column shifted sum into $X_1 \boxplus\!\!\rightarrow Y_1 = e_1 \otimes [A_k\ A_{k-1}\ \cdots\ A_0]$. In fact, this shifted sum operation is specifically designed to imitate the product of a pencil $L(\lambda) = \lambda X + Y$ with the block column matrix $\Lambda \otimes I_n$, in the sense of the following lemma.

Lemma 2.1.4. *Let* $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ *be an* $n \times n$ *matrix polynomial, and* $L(\lambda) = \lambda X + Y$ *a* $kn \times kn$ *pencil. Then for* $v \in \mathbb{F}^k$,

$$(\lambda X + Y) \cdot (\Lambda \otimes I_n) = v \otimes P(\lambda) \quad \Longleftrightarrow \quad X \boxplus\!\!\rightarrow Y = v \otimes [A_k \; A_{k-1} \; \cdots \; A_0], \tag{2.1.7}$$

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

$$\mathbb{L}_1(P) = \left\{ \lambda X + Y \; : \; X \boxplus\!\!\rightarrow Y = v \otimes [A_k \; A_{k-1} \; \cdots \; A_0], \; v \in \mathbb{F}^k \right\}. \tag{2.1.8}$$

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

Theorem 2.1.5 (Characterization of pencils in $\mathbb{L}_1(P)$).
Let $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ *be an* $n \times n$ *matrix polynomial, and* $v \in \mathbb{F}^k$ *any vector. Then the set of pencils in* $\mathbb{L}_1(P)$ *with right ansatz vector* v *consists of all* $L(\lambda) = \lambda X + Y$ *such that*

$$X = \begin{array}{c} \begin{array}{cc} \scriptstyle n & \scriptstyle (k-1)n \end{array} \\ \left[\begin{array}{cc} v \otimes A_k & -W \end{array} \right] \end{array} \quad \text{and} \quad Y = \begin{array}{c} \begin{array}{cc} \scriptstyle (k-1)n & \scriptstyle n \end{array} \\ \left[\begin{array}{cc} W + \big(v \otimes [\, A_{k-1} \; \cdots \; A_1 \,]\big) & v \otimes A_0 \end{array} \right] \end{array},$$

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

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

$$\begin{array}{rcl} \mathbb{L}_1(P) & \xrightarrow{\mathcal{M}} & \mathcal{V}_P \\ L(\lambda) & \longmapsto & L(\lambda)\,(\Lambda \otimes I_n) \,. \end{array} \tag{2.1.9}$$

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

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

Then $X_v \boxplus\!\!\rightarrow Y_v = v \otimes [A_k \; A_{k-1} \; \cdots \; A_0]$, so by Lemma 2.1.4 the pencil $L_v(\lambda) := \lambda X_v + Y_v$ is an $\mathcal{M}$-preimage of $v \otimes P(\lambda)$. The set of all $\mathcal{M}$-preimages of $v \otimes P(\lambda)$ is then $L_v(\lambda) + \ker \mathcal{M}$, so all that remains is to compute $\ker \mathcal{M}$. By (2.1.7), the kernel of $\mathcal{M}$ consists of all pencils $\lambda X + Y$ satisfying $X \boxplus\!\!\rightarrow Y = 0$. The definition of the shifted sum then implies that X and Y must have the form

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

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

Corollary 2.1.6. $\dim \mathbb{L}_1(P) = k(k-1)n^2 + k$ *.*

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

Thus we see that $\mathbb{L}_1(P)$ is a relatively large subspace of the full pencil space (with dimension $2k^2n^2$), yet the pencils in $\mathbb{L}_1(P)$ are still easy to construct from the data in P. The next corollary isolates a special case of Theorem 2.1.5 that plays an important role in Section 2.2.

Corollary 2.1.7. *Suppose $L(\lambda) = \lambda X + Y \in \mathbb{L}_1(P)$ has right ansatz vector $v = \alpha e_1$. Then*

$$X = \left[\begin{array}{c|c} \alpha A_k & X_{12} \\ \hline 0 & -Z \end{array}\right] \quad \text{and} \quad Y = \left[\begin{array}{c|c} Y_{11} & \alpha A_0 \\ \hline Z & 0 \end{array}\right] \tag{2.1.10}$$

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

Note that $C_1(\lambda)$ fits the pattern in Corollary 2.1.7 with $v = e_1$ and $Z = -I_{(k-1)n}$.

The second important property of the companion form is the simple relationship between its eigenvectors and those of the polynomial P that it linearizes. From the discussion following (2.1.2) it is evident that every eigenvector of $C_1(\lambda)$ has the form $\Lambda \otimes x$, where x is an eigenvector of P. Thus eigenvectors of P are recovered simply by extracting the last n coordinates from eigenvectors of the companion form. The next result shows that linearizations in $\mathbb{L}_1(P)$ also have this property.

Theorem 2.1.8 (Eigenvector Recovery Property for $\mathbb{L}_1(P)$).
Let $P(\lambda)$ be an $n \times n$ matrix polynomial of degree k, and $L(\lambda)$ any pencil in $\mathbb{L}_1(P)$ with nonzero right ansatz vector v. Then $x \in \mathbb{C}^n$ is an eigenvector for $P(\lambda)$ with finite eigenvalue $\lambda \in \mathbb{C}$ if and only if $\Lambda \otimes x$ is an eigenvector for $L(\lambda)$ with eigenvalue λ. If, in addition, P is regular and $L \in \mathbb{L}_1(P)$ is a linearization for P, then every eigenvector of L with finite eigenvalue λ is of the form $\Lambda \otimes x$ for some eigenvector x of P.

Proof. The first statement follows immediately from the identity

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

For the second statement, assume that $\lambda \in \mathbb{C}$ is a finite eigenvalue of $L(\lambda)$ with geometric multiplicity m, and let $y \in \mathbb{C}^{kn}$ be any eigenvector of $L(\lambda)$ associated with λ. Since $L(\lambda)$ is a linearization of $P(\lambda)$, the geometric multiplicity of λ for $P(\lambda)$ is also m. Let $x_1, \ldots, x_m$ be linearly independent eigenvectors of $P(\lambda)$ associated with λ, and define $y_i = \Lambda \otimes x_i$ for $i = 1, \ldots, m$. Then $y_1, \ldots, y_m$ are linearly independent eigenvectors for $L(\lambda)$ with eigenvalue λ, and so y must be a linear combination of $y_1, \ldots, y_m$. Thus y has the form $y = \Lambda \otimes x$ for some eigenvector $x \in \mathbb{C}^n$ for P. $\square$

A result analogous to Theorem 2.1.8 is also valid for the eigenvalue ∞. Because additional arguments are needed, this will be deferred until Section 2.2.

The above development and analysis of the pencil space $\mathbb{L}_1(P)$ has a parallel version in which the starting point is the second companion form $C_2(\lambda) = \lambda X_2 + Y_2$ as in (1.1.3). The analog of (2.1.4) is the identity

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

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

$$\big(\Lambda^T \otimes I_n\big) \cdot L(\lambda) = w^T \otimes P(\lambda), \tag{2.1.11}$$

and to a corresponding vector space $\mathbb{L}_2(P)$. The vector w in (2.1.11) will be referred to as the "left ansatz vector" for $L(\lambda)$.

Definition 2.1.9. With $\mathcal{W}_P = \{w^T \otimes P(\lambda) : w \in \mathbb{F}^k\}$, we define

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

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

Definition 2.1.10 (Row Shifted Sum). Let X and Y be block matrices

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

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

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

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

The following analog of Lemma 2.1.4 establishes the correspondence between the left ansatz and row shifted sums.

Lemma 2.1.11. *Let $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ be an $n \times n$ matrix polynomial, and $L(\lambda) = \lambda X + Y$ a $kn \times kn$ pencil. Then for any $w \in \mathbb{F}^k$,*

$$(\Lambda^T \otimes I_n) \cdot (\lambda X + Y) = w^T \otimes P(\lambda) \quad \Longleftrightarrow \quad X \boxplus\!\!\downarrow Y = w^T \otimes \begin{bmatrix} A_k \\ \vdots \\ A_0 \end{bmatrix}. \tag{2.1.12}$$

Using these tools one can characterize the pencils in $\mathbb{L}_2(P)$ in a manner completely analogous to Theorem 2.1.5, and thus conclude that

$$\dim \mathbb{L}_2(P) = \dim \mathbb{L}_1(P) = k(k-1)n^2 + k\,. \tag{2.1.13}$$

An even stronger relationship between the spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$, which again immediately implies (2.1.13), is established in Section 3.1 using the notion of block-transpose.

The analog of Theorem 2.1.8 for pencils in $\mathbb{L}_2(P)$ involves left eigenvectors of $P(\lambda)$ rather than right eigenvectors. Since the definition of a left eigenvector of a matrix polynomial does not seem to be completely standardized in the literature, we include here the definition used in this thesis.

Definition 2.1.12 (Left eigenvectors). A left eigenvector of an $n \times n$ matrix polynomial P associated with a finite eigenvalue λ is a nonzero vector $y \in \mathbb{C}^n$ such that $y^* P(\lambda) = 0$. A left eigenvector for P corresponding to the eigenvalue ∞ is a left eigenvector for $\operatorname{rev} P$ associated with the eigenvalue 0.

This definition differs from the one adopted in [31], although it is compatible with the usual definition for left eigenvectors of a matrix [33], [79]. We have chosen Definition 2.1.12 here because it is the one typically used in formulas for condition numbers of eigenvalues, a topic investigated in Chapter 8. The following result shows that left eigenvectors of P are easily recovered from linearizations in $\mathbb{L}_2(P)$. The proof is completely analogous to that given for Theorem 2.1.8.

Theorem 2.1.13 (Eigenvector Recovery Property for $\mathbb{L}_2(P)$). *Let $P(\lambda)$ be an $n\times n$ matrix polynomial of degree k, and $L(\lambda)$ any pencil in $\mathbb{L}_2(P)$ with nonzero left ansatz vector w. Then $y \in \mathbb{C}^n$ is a left eigenvector for $P(\lambda)$ with finite eigenvalue $\lambda \in \mathbb{C}$ if and only if $\overline{\Lambda} \otimes y$ is a left eigenvector for $L(\lambda)$ with eigenvalue λ. If, in addition, P is regular and $L \in \mathbb{L}_2(P)$ is a linearization for P, then every left eigenvector of L with finite eigenvalue λ is of the form $\overline{\Lambda} \otimes y$ for some left eigenvector y of P.*

Just as for Theorem 2.1.8, there is an analogous result for the eigenvalue ∞ that can be found in Section 2.2.

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

2.2 When is a Pencil in $\mathbb{L}_1(P)$ a Linearization?

It is clear that not all pencils in the spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ are linearizations of P — consider, for example, any pencil in $\mathbb{L}_1(P)$ with right ansatz vector $v = 0$. In this section we focus on $\mathbb{L}_1(P)$ and obtain criteria for deciding whether a pencil from $\mathbb{L}_1(P)$ is a linearization for P or not. We show, for example, that for any given $L \in \mathbb{L}_1(P)$ there is typically a condition (specific to L) on the coefficient matrices of P that must be satisfied in order to guarantee that L is actually a linearization for P. Specific examples of such "linearization conditions" can be found in Section 2.2.2 and in the tables in Chapter 4. Analogs of all the results in this section also hold for $\mathbb{L}_2(P)$, with very similar arguments.

2.2.1 The Strong Linearization Theorem

We begin with a result concerning the special case of the right ansatz (2.1.5) considered in Corollary 2.1.7. Note that P is not assumed here to be regular.

Theorem 2.2.1. *Suppose $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ with $A_k \neq 0$ is an $n \times n$ matrix polynomial, and $L(\lambda) = \lambda X + Y \in \mathbb{L}_1(P)$ has nonzero right ansatz vector $v = \alpha e_1$, so that*

$$L(\lambda) \cdot (\Lambda \otimes I_n) = \alpha e_1 \otimes P(\lambda)\,. \tag{2.2.1}$$

Partition X and Y as in (2.1.10) *so that*

$$L(\lambda) \;=\; \lambda X + Y \;=\; \lambda \left[\begin{array}{c|c} \alpha A_k & X_{12} \\ \hline 0 & -Z \end{array}\right] + \left[\begin{array}{c|c} Y_{11} & \alpha A_0 \\ \hline Z & 0 \end{array}\right], \tag{2.2.2}$$

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

Proof. We show first that $L(\lambda)$ is a linearization of $P(\lambda)$. Begin the reduction of $L(\lambda)$ to $\mathrm{diag}(P(\lambda), I_{(k-1)n})$ using the unimodular matrix polynomials $T(\lambda)$ and $G(\lambda)$ defined in (1.1.6). In the product $L(\lambda)G(\lambda)$, clearly the first $k-1$ block-columns are the same as those of $L(\lambda)$; because the last block-column of $G(\lambda)$ is $\Lambda \otimes I$, we see from (2.2.1) that the last block-column of $L(\lambda)G(\lambda)$ is $\alpha e_1 \otimes P(\lambda)$. Partitioning Z in (2.2.2) into block columns $[Z_1\ Z_2\ \dots\ Z_{k-1}]$, where $Z_i \in \mathbb{F}^{(k-1)n\times n}$, we thus obtain

$$\begin{aligned} L(\lambda)G(\lambda) &= \begin{bmatrix} * & * & \dots & * & * \\ Z_1 & (Z_2-\lambda Z_1) & \dots & (Z_{k-1}-\lambda Z_{k-2}) & -\lambda Z_{k-1} \end{bmatrix} G(\lambda)\,, \\ &= \begin{bmatrix} * & * & \dots & * & \alpha P(\lambda) \\ Z_1 & (Z_2-\lambda Z_1) & \dots & (Z_{k-1}-\lambda Z_{k-2}) & 0 \end{bmatrix}. \end{aligned}$$

Further transformation by block-column operations yields

$$L(\lambda)T(\lambda) = L(\lambda)\underbrace{G(\lambda)\begin{bmatrix} I & \lambda I & & & \\ & I & & & \\ & & I & & \\ & & & \ddots & \\ & & & & I \end{bmatrix}\begin{bmatrix} I & & & & \\ & I & \lambda I & & \\ & & I & & \\ & & & \ddots & \\ & & & & I \end{bmatrix}\cdots\begin{bmatrix} I & & & \\ & \ddots & & \\ & & I & \lambda I \\ & & & I \end{bmatrix}}_{=T(\lambda)} = \left[\begin{array}{c|c} * & \alpha P(\lambda) \\ \hline Z & 0 \end{array}\right].$$

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

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

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

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

To show that $L(\lambda)$ is also a strong linearization for $P(\lambda)$, it remains to show that $\mathrm{rev}\,L(\lambda) = \lambda Y + X$ is a linearization for $\mathrm{rev}\,P(\lambda)$. Now it would be nice if $\mathrm{rev}\,L(\lambda)$ was a pencil in $\mathbb{L}_1(\mathrm{rev}P)$, but it is not; however, a small modification of $\mathrm{rev}\,L(\lambda)$ is in $\mathbb{L}_1(\mathrm{rev}P)$. Observe that $\lambda^{k-1}\cdot\Lambda(1/\lambda) = [1, \lambda, \dots, \lambda^{k-2}, \lambda^{k-1}]^T = R_k\Lambda$, where R_k denotes the $k \times k$ reverse identity matrix. Thus replacing λ by $1/\lambda$ in (2.2.1) and multiplying both sides by λ^k yields

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

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

$$\widehat{L}(\lambda)\cdot(\Lambda\otimes I) = \alpha e_1 \otimes \mathrm{rev}\,P(\lambda), \tag{2.2.3}$$

and so $\widehat{L} \in \mathbb{L}_1(\operatorname{rev} P)$. (Observe that $\widehat{L}(\lambda)$ is just $\operatorname{rev} L(\lambda) = \lambda Y + X$ with the block-columns of Y and X arranged in reverse order.) Since $\widehat{L}$ and $\operatorname{rev} L$ are equivalent pencils, the proof will be complete once we show that $\lambda\widehat{X}+\widehat{Y} := \widehat{L}(\lambda)$ is a linearization for $\operatorname{rev} P(\lambda)$. But $\widehat{X} = Y \cdot (R_k \otimes I)$ and $\widehat{Y} = X \cdot (R_k \otimes I)$, and hence from (2.2.2) it follows that

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

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

Remark 2.2.2. *The fact (first proved in [30]) that the first companion form of any polynomial is always a strong linearization is a special case of Theorem* 2.2.1.

When a matrix polynomial $P(\lambda)$ is regular, then it is easy to see from Definition 1.1.4 that any linearization for $P(\lambda)$ must also be regular. The next result shows something rather surprising: when a pencil L is in $\mathbb{L}_1(P)$, the minimal necessary condition of regularity is actually sufficient to guarantee that L is a linearization for P. This serves to emphasize just how close a pencil is to being a linearization for P, even a strong linearization for P, once it satisfies the ansatz (2.1.5).

Theorem 2.2.3 (Strong Linearization Theorem).
Let $L(\lambda) \in \mathbb{L}_1(P)$ for a regular matrix polynomial $P(\lambda)$. Then the following statements are equivalent:

(i) $L(\lambda)$ is a linearization for $P(\lambda)$.

(ii) $L(\lambda)$ is a regular pencil.

(iii) $L(\lambda)$ is a strong linearization for $P(\lambda)$.

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

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

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

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

Hence by Corollary 2.1.7 the matrices $\widetilde{X}$ and $\widetilde{Y}$ in $\widetilde{L}(\lambda) := \lambda\widetilde{X} + \widetilde{Y}$ have the forms

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

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

$$\left[\begin{array}{cc} 0 & w^T \end{array}\right](\lambda\widetilde{X} + \widetilde{Y}) = 0 \quad \text{for all } \lambda \in \mathbb{F},$$

contradicting the regularity of $\widetilde{L}(\lambda)$. Thus $\widetilde{Z}$ is nonsingular, and so by Theorem 2.2.1 we know that $\widetilde{L}(\lambda)$, and hence also $L(\lambda)$, is a strong linearization for $P(\lambda)$.

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

Now recall from Definitions 1.1.3 and 2.1.12 that a vector $x \in \mathbb{C}^n$ is a right (left) eigenvector for a polynomial P with eigenvalue ∞ if and only if x is a right (left) eigenvector for $\mathrm{rev}P$ with eigenvalue 0. Translating statements about infinite eigenvalues to ones about zero eigenvalues allows us to use Theorems 2.1.8, 2.1.13, and 2.2.3 to extend the eigenvector recovery properties of $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ to the eigenvalue ∞.

Theorem 2.2.4 (Eigenvector Recovery at ∞).
Let $P(\lambda)$ be an $n \times n$ matrix polynomial of degree k, and $L(\lambda)$ any pencil in $\mathbb{L}_1(P)$ (resp., $\mathbb{L}_2(P)$) with nonzero right (left) ansatz vector v. Then $x \in \mathbb{C}^n$ is a right (left) eigenvector for $P(\lambda)$ with eigenvalue ∞ if and only if $e_1 \otimes x$ is a right (left) eigenvector for $L(\lambda)$ with eigenvalue ∞. If, in addition, P is regular and $L \in \mathbb{L}_1(P)$ (resp., $\mathbb{L}_2(P)$) is a linearization for P, then every right (left) eigenvector of L with eigenvalue ∞ is of the form $e_1 \otimes x$ for some right (left) eigenvector x of P with eigenvalue ∞.

Proof. We give the proof only for right eigenvectors of $L \in \mathbb{L}_1(P)$ here. The argument for recovery of left eigenvectors of $L \in \mathbb{L}_2(P)$ is essentially the same, given the analogs of Theorems 2.2.1 and 2.2.3 for $\mathbb{L}_2(P)$.

For any $L(\lambda)$ define $\widehat{L}(\lambda) := \mathrm{rev}L(\lambda) \cdot (R_k \otimes I)$. Then the reasoning used in Theorem 2.2.1 to obtain (2.2.3) shows that $L \in \mathbb{L}_1(P) \Rightarrow \widehat{L} \in \mathbb{L}_1(\mathrm{rev}P)$, with the same nonzero right ansatz vector v. By Theorem 2.1.8 we know that x is a right eigenvector for $\mathrm{rev}P$ with eigenvalue 0 if and only if $\Lambda \otimes x = e_k \otimes x$ is a right eigenvector for $\widehat{L}$ with eigenvalue 0. But $e_k \otimes x$ is a right eigenvector for $\widehat{L}$ if and only if $e_1 \otimes x = (R_k \otimes I)(e_k \otimes x)$ is a right eigenvector for $\mathrm{rev}L$, both with eigenvalue 0. This establishes the first part of the theorem.

If P is regular and $L \in \mathbb{L}_1(P)$ is a linearization for P, then by Theorem 2.2.3 $\widehat{L} \in \mathbb{L}_1(\mathrm{rev}P)$ is a linearization for $\mathrm{rev}P$. Theorem 2.1.8 then implies that every right eigenvector of $\widehat{L}$ with eigenvalue 0 is of the form $e_k \otimes x$, where x is a right eigenvector of $\mathrm{rev}P$ with eigenvalue 0; equivalently every right eigenvector of $\mathrm{rev}L$ with eigenvalue 0 is of the form $e_1 \otimes x$ for some right eigenvector x of $\mathrm{rev}P$ with eigenvalue 0. This establishes the second part of the theorem. ∎

2.2.2 Linearization Conditions

A useful by-product of the proof of Theorem 2.2.3 is a simple procedure for generating a symbolic "linearization condition" for any given pencil $L \in \mathbb{L}_1(P)$, i.e., a necessary and sufficient condition (in terms of the data in P) for L to be a linearization for P. We describe this procedure and then illustrate with some examples.

Procedure to determine the linearization condition for a pencil in $\mathbb{L}_1(P)$:

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

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

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

$$\widetilde{L}(\lambda) = \lambda \widetilde{X} + \widetilde{Y} = \lambda \left[\begin{array}{c|c} \widetilde{X}_{11} & \widetilde{X}_{12} \\ \hline 0 & -Z \end{array}\right] + \left[\begin{array}{c|c} \widetilde{Y}_{11} & \widetilde{Y}_{12} \\ \hline Z & 0 \end{array}\right], \tag{2.2.4}$$

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

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

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

Example 2.2.5. *Consider the general quadratic polynomial $P(\lambda) = \lambda^2 A + \lambda B + C$ (assumed to be regular) and the following pencils in $\mathbb{L}_1(P)$:*

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

Since

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

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

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

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

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

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

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

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

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

Example 2.2.6. *Consider the general cubic polynomial* $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$ (*again assumed to be regular*) *and the pencil*

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

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

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

and hence the linearization condition $\det Z = \det(B + C - DA) \neq 0$. (*Recall that for* $n \times n$ *blocks* W, X, Y, Z *with* $YZ = ZY$, *we have* $\det \left[\begin{smallmatrix} W & X \\ Y & Z \end{smallmatrix}\right] = \det(WZ - XY)$. *See [56].*)

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

Theorem 2.2.7 (Linearizations are Generic in $\mathbb{L}_1(P)$).
For any regular $n \times n$ *matrix polynomial* $P(\lambda)$ *of degree* k, *almost every pencil in* $\mathbb{L}_1(P)$ *is a linearization for* $P(\lambda)$. (*Here by "almost every" we mean for all but a closed, nowhere dense set of measure zero in* $\mathbb{L}_1(P)$.)

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

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

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

Now by Theorem 2.2.3 we know that $L(\lambda) \in \mathbb{L}_1(P)$ *fails* to be a linearization for $P(\lambda)$ if and only if $\det L(\lambda) \equiv 0$, equivalently if all the coefficients c_i are zero. Thus the subset of pencils in $\mathbb{L}_1(P)$ that are not linearizations for $P(\lambda)$ can be characterized as the common zero set $\mathcal{Z}$ of the polynomials $\{c_i(\beta_1, \beta_2, \ldots, \beta_d) : 0 \leq i \leq kn\}$, i.e., as an algebraic subset of $\mathbb{F}^d$.

Since *proper* algebraic subsets of $\mathbb{F}^d$ are well known to be closed, nowhere dense subsets of measure zero, the proof will be complete once we show that $\mathcal{Z}$ is a proper subset of $\mathbb{F}^d$, or equivalently, that there exists some pencil in $\mathbb{L}_1(P)$ that is a linearization for P. But this is immediate: the first companion form $C_1(\lambda)$ for $P(\lambda)$ is in $\mathbb{L}_1(P)$ and is always a linearization for P (see [31] or Remark 2.2.2). $\square$

Although $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ contain a large supply of linearizations for P, there do exist simple linearizations for P that are neither in $\mathbb{L}_1(P)$ nor in $\mathbb{L}_2(P)$. We illustrate this with a recent example from [3].

Example 2.2.8. *For the cubic matrix polynomial $P(\lambda) = \lambda^3 A_3 + \lambda^2 A_2 + \lambda A_1 + A_0$, the pencil*

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

is shown in [3] to be a linearization for P. Using shifted sums it is easy to see that $L(\lambda)$ is in neither $\mathbb{L}_1(P)$ nor $\mathbb{L}_2(P)$.

2.3 Another View of $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$

In Section 2.1 we defined the pencil space $\mathbb{L}_1(P)$ by generalizing one particular property of the first companion form $C_1(\lambda)$ of P. A different connection between $\mathbb{L}_1(P)$ and $C_1(\lambda)$ can be established, which gives an alternative insight into why the pencils in $\mathbb{L}_1(P)$ retain so many of the attractive features of $C_1(\lambda)$. Using the first three steps of the procedure in Section 2.2.2, together with the characterization of $\mathbb{L}_1(P)$-pencils given in Theorem 2.1.5 and Corollary 2.1.7, one can show that any $L(\lambda) \in \mathbb{L}_1(P)$ can be factored (non-uniquely) in the form

$$L(\lambda) \;=\; (K \otimes I_n) \left[\begin{array}{c|c} \alpha I_n & U \\ \hline 0 & -Z \end{array}\right] C_1(\lambda)\,, \tag{2.3.1}$$

where $Z \in \mathbb{F}^{(k-1)n \times (k-1)n}$ is the same as the block Z in Corollary 2.1.7 and (2.2.4), and $K \in \mathbb{F}^{k \times k}$ is nonsingular. It is also straightforward to check that any pencil of the form (2.3.1) satisfies $L(\lambda) \cdot (\Lambda \otimes I_n) = \alpha K e_1 \otimes P(\lambda)$, and so is in $\mathbb{L}_1(P)$ with right ansatz vector $v = \alpha K e_1$. Consequently the scalar $\alpha \in \mathbb{F}$ in the middle factor is zero if and only if the right ansatz vector v of $L(\lambda)$ is zero. This factorization gives another reason why the *right* eigenvectors of pencils in $\mathbb{L}_1(P)$ have the same Kronecker product structure as those of $C_1(\lambda)$, and why pencils in $\mathbb{L}_1(P)$ are either strong linearizations of P (like $C_1(\lambda)$) or singular pencils, depending on the nonsingularity or singularity of the block Z and the scalar α.

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

$$L(\lambda) \;=\; C_2(\lambda) \left[\begin{array}{c|c} \beta I_n & 0 \\ \hline T & -V \end{array}\right] (H \otimes I_n)\,, \tag{2.3.2}$$

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

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

We are also interested in the possibility of the existence of pencils that are simultaneously in $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. The factored forms (2.3.1) and (2.3.2) might make it seem rather unlikely that there could be any nontrivial pencils in this intersection. However, in the next chapter we will see (using shifted sums) that this is an erroneous impression.

Finally, it is worth pointing out that the ansatz equations (2.1.5) and (2.1.11) enjoy the advantage of being *identities* in the variable λ, and so can be treated analytically as well as algebraically. This property is exploited in the analysis of the conditioning of eigenvalues of linearizations in Chapter 8.

Chapter 3

$\mathbb{DL}(P)$ and Block-symmetry

The spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ provide large sources of linearizations for any regular matrix polynomial P, structured or unstructured, but pencils in the *intersection* $\mathbb{L}_1(P) \cap \mathbb{L}_2(P)$ are of particular interest. Any such pencil has simultaneously both left and right eigenvector recovery properties as described in Theorems 2.1.8, 2.1.13 and 2.2.4, and so is especially amenable to an analysis of the conditioning of its eigenvalues, as will be seen later in Chapter 8. Hence we introduce some special notation for this intersection.

Definition 3.0.1. $\mathbb{DL}(P) := \mathbb{L}_1(P) \cap \mathbb{L}_2(P)$.

The primary goal of this chapter is to understand the space $\mathbb{DL}(P)$ of "double ansatz" pencils for a *general* matrix polynomial P, and to show how all the pencils in $\mathbb{DL}(P)$ may be explicitly constructed. A priori there is no obvious reason why $\mathbb{DL}(P)$ should contain any nontrivial pencils at all. However, initial investigation of these pencil spaces for P of degree 2 and 3 indicates that nontrivial $\mathbb{DL}(P)$-pencils do exist, and that they all seem to have an unanticipated additional structure — *block-symmetry* — even when P is itself *un*structured. It turns out that a preliminary independent study of block-symmetric pencils in $\mathbb{L}_1(P)$ provides an efficient pathway to the properties of $\mathbb{DL}(P)$; hence we begin this chapter with the notions of block-transpose and block-symmetry, and a development of some of their basic properties.

3.1 The Block-transpose Operation

This section introduces the simple operation of *block-transpose*, and shows how it establishes a fundamental relationship between the spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$.

Recall from chapter 2 that pencils in $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ are of size $kn \times kn$, where k is the degree of the $n \times n$ matrix polynomial $P(\lambda)$. The shifted sum operations $\boxplus\!\!\rightarrow$ and $\underset{\downarrow}{\boxplus}$ treat these pencils as block $k \times k$ matrices with blocks of size $n \times n$, and this convention concerning pencils in $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ will be maintained throughout. Indeed, for the purposes of this thesis we only consider block matrices in which all the blocks have the same size, as in the following formal definition.

Definition 3.1.1 (Block matrix). Let $E_{ij} \in \mathbb{F}^{k\times\ell}$ denote the matrix that is everywhere zero except for a 1 in the (i, j) entry. Then a $km \times \ell n$ matrix A written in the form

$$A = \sum_{i,j} (E_{ij} \otimes B_{ij})$$

with $B_{ij} \in \mathbb{F}^{m\times n}$ is said to be a *block $k \times \ell$ matrix* with $m \times n$ blocks B_{ij}.

For this restricted class of block matrices one can give a very simple and straightforward description of the block-transpose operation. Let $A = (B_{ij})$ be a block $k \times \ell$ matrix with $m \times n$ blocks B_{ij}. Then the block transpose of A is the block $\ell \times k$ matrix $A^{\mathcal{B}}$, with $m \times n$ blocks defined by $(A^{\mathcal{B}})_{ij} = B_{ji}$. However, for giving formal proofs of some of the basic properties of block-transpose the following definition is sometimes more convenient.

Definition 3.1.2 (Block-transpose). Suppose $A = \sum_{i,j}(E_{ij} \otimes B_{ij})$ is a block $k \times \ell$ matrix with $m \times n$ blocks B_{ij}. Then the *block-transpose* of A is the block $\ell \times k$ matrix

$$A^{\mathcal{B}} := \sum_{i,j} (E_{ij}^T \otimes B_{ij}),$$

also with $m \times n$ blocks. Note that if A is expressed as a sum $A = \sum (K_i \otimes M_i)$ where $K_i \in \mathbb{F}^{k\times\ell}$ and $M_i \in \mathbb{F}^{m\times n}$, then

$$A^{\mathcal{B}} = \sum (K_i^T \otimes M_i).$$

Block-transpose has some, but not all, of the same algebraic properties as ordinary matrix transpose.

Lemma 3.1.3. *For any block matrices A, C that are conformable for block-addition we have*

$$(A + C)^{\mathcal{B}} \;=\; A^{\mathcal{B}} + C^{\mathcal{B}}. \tag{3.1.1}$$

Proof. The straightforward proof is omitted. ∎

Note that block-matrices A and C do *not* in general satisfy $(AC)^{\mathcal{B}} = C^{\mathcal{B}}A^{\mathcal{B}}$. However, if each block of A commutes with each block of C then the analog of the multiplicative property of transpose does hold.

Lemma 3.1.4. *Suppose $A = \sum_{i,j}(E_{ij} \otimes A_{ij})$ is a $(k \times \ell)$-block-matrix with $n \times n$ blocks A_{ij}, and $C = \sum_{p,q}(E_{pq} \otimes C_{pq})$ is an $(\ell \times m)$-block-matrix with $n \times n$ blocks C_{pq}. Further suppose that each A_{ij} commutes with each C_{pq}. Then*

$$(AC)^{\mathcal{B}} \;=\; C^{\mathcal{B}}A^{\mathcal{B}}. \tag{3.1.2}$$

Proof. We have

$$
\begin{aligned}
(AC)^{\mathcal{B}} &= \left(\sum_{i,j}(E_{ij}\otimes A_{ij}) \cdot \sum_{p,q}(E_{pq}\otimes C_{pq})\right)^{\mathcal{B}} \\
&= \left(\sum_{i,j,p,q}(E_{ij}E_{pq}\otimes A_{ij}C_{pq})\right)^{\mathcal{B}} \\
&= \sum_{i,j,p,q}\Big((E_{ij}E_{pq})^T\otimes A_{ij}C_{pq}\Big) \\
&= \sum_{i,j,p,q}\big(E_{pq}^T E_{ij}^T\otimes C_{pq}A_{ij}\big) \\
&= \sum_{p,q}\big(E_{pq}^T\otimes C_{pq}\big) \cdot \sum_{i,j}\big(E_{ij}^T\otimes A_{ij}\big) \;=\; C^{\mathcal{B}}A^{\mathcal{B}}. \qquad \square
\end{aligned}
$$

Remark 3.1.5. Note that the hypothesis of A_{ij} commuting with every C_{pq} is stronger than needed in Lemma 3.1.4. It suffices to assume that A_{ij} commutes with every C_{pq} with $p = j$ for the calculation proving Lemma 3.1.4 to be valid.

The block-transpose operation establishes an intimate link between the pencil spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. Keep in mind our convention that pencils in $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ are to be regarded as block $k \times k$ matrices with $n \times n$ blocks, and that block-transpose is performed relative to this partitioning.

Theorem 3.1.6 (A Fundamental Isomorphism).
For any matrix polynomial $P(\lambda)$, the block-transpose map

$$
\begin{aligned}
\mathbb{L}_1(P) &\xrightarrow{\;\mathfrak{B}\;} \mathbb{L}_2(P) \\
L(\lambda) &\longmapsto L(\lambda)^{\mathcal{B}}
\end{aligned}
$$

is a linear isomorphism between $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$. In particular, if $L(\lambda) \in \mathbb{L}_1(P)$ has right ansatz vector v, then $L(\lambda)^{\mathcal{B}} \in \mathbb{L}_2(P)$ with left ansatz vector $w = v$.

Proof. Suppose $L(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector v. Then using Definition 3.1.2 and Lemma 3.1.4 we have

$$
\begin{aligned}
\big(L(\lambda)\cdot(\Lambda\otimes I)\big)^{\mathcal{B}} = \big(v\otimes P(\lambda)\big)^{\mathcal{B}} \quad &\Longrightarrow \quad (\Lambda\otimes I)^{\mathcal{B}}\cdot L(\lambda)^{\mathcal{B}} \;=\; v^T\otimes P(\lambda) \\
&\Longrightarrow \quad (\Lambda^T\otimes I)\cdot L(\lambda)^{\mathcal{B}} \;=\; v^T\otimes P(\lambda)\,.
\end{aligned}
$$

This last equation says that $L(\lambda)^{\mathcal{B}} \in \mathbb{L}_2(P)$ with left ansatz vector v, and so block-transpose gives a well-defined map $\mathfrak{B}$ from $\mathbb{L}_1(P)$ to $\mathbb{L}_2(P)$. By Lemma 3.1.3 this map is linear. The kernel is clearly just the zero pencil, since $L(\lambda)^{\mathcal{B}} = 0 \Rightarrow L(\lambda) = 0$, so $\mathfrak{B}$ is injective. An analogous argument shows that block-transpose also gives a well-defined one-to-one linear map *backwards* from $\mathbb{L}_2(P)$ to $\mathbb{L}_1(P)$, thus showing $\mathfrak{B}$ to be an isomorphism. $\square$

Remark 3.1.7. In section 2.1 it was shown that $\dim \mathbb{L}_1(P) = \dim \mathbb{L}_2(P)$. Note that Theorem 3.1.6 provides an *independent* proof of that fact.

Example 3.1.8. The companion forms give a nice illustration of Theorem 3.1.6. One sees by inspection that $C_2(\lambda) = \big(C_1(\lambda)\big)^{\mathcal{B}}$; observe also that $C_1(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector $v = e_1$ while $C_2(\lambda) \in \mathbb{L}_2(P)$ with left ansatz vector $w = v = e_1$.

The final result of this section shows how the block-transpose operation interacts with shifted sums, and provides a different insight into how block-transpose connects $\mathbb{L}_1(P)$ with $\mathbb{L}_2(P)$.

Lemma 3.1.9. *For any block $k \times k$ matrices X and Y (with $n \times n$ blocks) we have*

$$(X \boxplus\!\!\rightarrow Y)^{\mathcal{B}} = X^{\mathcal{B}} \boxplus\!\!\!\downarrow Y^{\mathcal{B}}. \tag{3.1.3}$$

Proof. By definition of the row and column shifted sums,

$$\begin{aligned}(X \boxplus\!\!\rightarrow Y)^{\mathcal{B}} = \big(\left[\begin{array}{c|c} X & 0 \end{array}\right] + \left[\begin{array}{c|c} 0 & Y \end{array}\right]\big)^{\mathcal{B}} &= \left[\begin{array}{c|c} X & 0 \end{array}\right]^{\mathcal{B}} + \left[\begin{array}{c|c} 0 & Y \end{array}\right]^{\mathcal{B}} \\ &= \left[\begin{array}{c} X^{\mathcal{B}} \\ \hline 0 \end{array}\right] + \left[\begin{array}{c} 0 \\ \hline Y^{\mathcal{B}} \end{array}\right] = X^{\mathcal{B}} \boxplus\!\!\!\downarrow Y^{\mathcal{B}}. \qquad \square\end{aligned}$$

Remark 3.1.10. Note that an alternative proof of Theorem 3.1.6 can be given using the shifted sum property (3.1.3). If $L(\lambda) = \lambda X + Y \in \mathbb{L}_1(P)$ with right ansatz vector v, then $X \boxplus\!\!\rightarrow Y = v \otimes [A_k \; A_{k-1} \; \cdots \; A_0]$; from Lemma 3.1.9 we have

$$\big(X \boxplus\!\!\rightarrow Y\big)^{\mathcal{B}} = \big(v \otimes [A_k \; A_{k-1} \; \cdots \; A_0]\big)^{\mathcal{B}} \quad \Longrightarrow \quad X^{\mathcal{B}} \boxplus\!\!\!\downarrow Y^{\mathcal{B}} = v^T \otimes \begin{bmatrix} A_k \\ \vdots \\ A_0 \end{bmatrix}.$$

Thus $L(\lambda)^{\mathcal{B}} = \lambda X^{\mathcal{B}} + Y^{\mathcal{B}}$ is an element of $\mathbb{L}_2(P)$ with left ansatz vector v. The rest of the proof then proceeds as before.

3.2 Block-symmetry and Shifted Sum Equations

Given the notion of block-transpose, it is natural to consider block-symmetric matrices, which will play a central role in our development.

Definition 3.2.1 (Block-symmetry). A block $k \times k$ matrix A with $m \times n$ blocks is said to be *block-symmetric* if $A^{\mathcal{B}} = A$.

For example, a block 2×2 matrix is block-symmetric if and only if it has the form $\left[\begin{smallmatrix} A & C \\ C & D \end{smallmatrix}\right]$. Note also that if each block $A_{ij} \in \mathbb{F}^{n\times n}$ in a block-symmetric matrix A is symmetric, then A is itself symmetric.

It is easy to see that the set of all block-symmetric matrices of any fixed shape is closed under linear combinations, and hence forms a vector space. Thus we consider

$$\mathbb{B}(P) := \big\{\, \lambda X + Y \in \mathbb{L}_1(P) : X^{\mathcal{B}} = X,\; Y^{\mathcal{B}} = Y \,\big\} \subseteq \mathbb{L}_1(P), \tag{3.2.1}$$

the subspace of all block-symmetric pencils in $\mathbb{L}_1(P)$. As an immediate corollary of Theorem 3.1.6 we see that any block-symmetric pencil in $\mathbb{L}_1(P)$ is automatically in $\mathbb{DL}(P)$. This fact motivates a thorough preliminary study of $\mathbb{B}(P)$ as preparation for a characterization of $\mathbb{DL}(P)$.

Corollary 3.2.2. *For any matrix polynomial $P(\lambda)$, $\mathbb{B}(P) \subseteq \mathbb{DL}(P)$.*

Proof. Let $L(\lambda) \in \mathbb{B}(P) \subset \mathbb{L}_1(P)$. From Theorem 3.1.6 we know that $L(\lambda)^{\mathcal{B}} = L(\lambda)$ is in $\mathbb{L}_2(P)$, and so $L(\lambda) \in \mathbb{DL}(P)$. ∎

Because of Lemma 2.1.4, the construction of pencils in $\mathbb{B}(P)$ reduces to finding solutions of the shifted sum equation $X \boxplus\!\!\rightarrow Y = Z$ with block-symmetric X and Y. It is to this task that we turn next.

3.2.1 Shifted Sum Equations

We investigate the existence and uniqueness of block-symmetric solutions of the equation $X \boxplus\!\!\rightarrow Y = Z$, where Z is a given arbitrary matrix. We show that the equation $X \boxplus\!\!\rightarrow Y = Z$ may *always* be solved with block-symmetric X and Y, and that the *only* block-symmetric solution of $X \boxplus\!\!\rightarrow Y = 0$ is $X = Y = 0$.

First we define several special types of block-symmetric matrix that play a central role in the constructions to come. Let

$$R_\ell = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}_{\ell\times\ell} \quad \text{and} \quad N_\ell = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}_{\ell\times\ell}. \quad \text{(Note that } N_1 = [\,0\,].) \tag{3.2.2}$$

Then for an arbitrary $n \times n$ block M, define three block-Hankel, block-symmetric, block $\ell \times \ell$ matrices as follows:

$$\begin{aligned}
\mathcal{H}_\ell^{(0)}(M) &:= \begin{bmatrix} & & M \\ & \iddots & \\ M & & \end{bmatrix} = R_\ell \otimes M, \\
\mathcal{H}_\ell^{(1)}(M) &:= \begin{bmatrix} & & M & 0 \\ & \iddots & \iddots & \\ M & \iddots & & \\ 0 & & & \end{bmatrix} = (N_\ell R_\ell) \otimes M = \begin{bmatrix} & & 1 & 0 \\ & \iddots & \iddots & \\ 1 & \iddots & & \\ 0 & & & \end{bmatrix} \otimes M, \\
\mathcal{H}_\ell^{(-1)}(M) &:= \begin{bmatrix} & & & 0 \\ & & \iddots & M \\ & \iddots & \iddots & \\ 0 & M & & \end{bmatrix} = (R_\ell N_\ell) \otimes M = \begin{bmatrix} & & & 0 \\ & & \iddots & 1 \\ & \iddots & \iddots & \\ 0 & 1 & & \end{bmatrix} \otimes M.
\end{aligned}$$

The superscript (0), (1), or (-1) here specifies that the blocks M are on, above, or below the anti-diagonal, respectively. Note that all three of these block-Hankel matrices are symmetric if M is.

Lemma 3.2.3. *Let Z be an arbitrary block $k \times (k+1)$ matrix with $n \times n$ blocks. Then there exist block-symmetric block $k \times k$ matrices X and Y with $n \times n$ blocks such that $X \boxplus\!\!\rightarrow Y = Z$.*

Proof. Let $E_{ij}^\ell \in \mathbb{F}^{\ell\times(\ell+1)}$ denote the matrix that is everywhere zero except for a 1 in the (i,j) entry. Our proof is based on the observation that for arbitrary $M, P \in \mathbb{F}^{n\times n}$, the shifted sums

$$\mathcal{H}_\ell^{(0)}(M) \boxplus\!\!\rightarrow \left(-\mathcal{H}_\ell^{(1)}(M)\right) = \begin{bmatrix} 0 & \dots & \dots & 0 & 0 \\ \vdots & & \iddots & \iddots & \vdots \\ 0 & 0 & \iddots & & \vdots \\ M & 0 & \dots & \dots & 0 \end{bmatrix} = E_{\ell 1}^\ell \otimes M, \tag{3.2.3}$$

$$-\mathcal{H}_\ell^{(-1)}(P) \boxplus\!\!\rightarrow \mathcal{H}_\ell^{(0)}(P) = \begin{bmatrix} 0 & \dots & \dots & 0 & P \\ \vdots & & \iddots & 0 & 0 \\ \vdots & \iddots & \iddots & & \vdots \\ 0 & 0 & \dots & \dots & 0 \end{bmatrix} = E^{\ell}_{1,\ell+1} \otimes P \tag{3.2.4}$$

place M and P at the bottom left corner and top right corner of a block $\ell \times (\ell+1)$ matrix, respectively.

The shifted sum $\boxplus\!\!\rightarrow$ is compatible with ordinary sums, i.e.,

$$\left(\sum X_i\right) \boxplus\!\!\rightarrow \left(\sum Y_i\right) = \sum (X_i \boxplus\!\!\rightarrow Y_i).$$

Hence if we can show how to construct block-symmetric X and Y that place an arbitrary $n \times n$ block into an arbitrary (i,j) block-location in Z, then sums of such examples will achieve the desired result for an arbitrary Z.

For indices i, j such that $1 \le i \le j \le k$, let $\ell = j - i + 1$ and embed $\mathcal{H}_\ell^{(0)}(M)$ and $-\mathcal{H}_\ell^{(1)}(M)$ as a principal submatrix in block rows and block columns i through j of the block $k \times k$ zero matrix to get

$$\begin{aligned}
\widetilde{X}_{ij} \boxplus\!\!\rightarrow \widetilde{Y}_{ij} \;:=\; & \begin{array}{c c} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ j \end{array} & \begin{bmatrix} & & \\ & \boxed{\mathcal{H}_\ell^{(0)}(M)} & \\ & & \end{bmatrix} \end{array} \boxplus\!\!\rightarrow \begin{array}{c c} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ j \end{array} & \begin{bmatrix} & & \\ & \boxed{-\mathcal{H}_\ell^{(1)}(M)} & \\ & & \end{bmatrix} \end{array} \qquad (3.2.5) \\
= \;& \begin{array}{c c} & \begin{array}{cc} i & j+1 \end{array} \\ \begin{array}{c} i \\ j \end{array} & \begin{bmatrix} & & \\ & \boxed{\mathcal{H}_\ell^{(0)}(M) \boxplus\!\!\rightarrow (-\mathcal{H}_\ell^{(1)}(M))} & \\ & & \end{bmatrix} \end{array} \\
= \;& E_{ji} \otimes M \quad (i \le j).
\end{aligned}$$

Note that embedding $\mathcal{H}_\ell^{(0)}(M)$ and $-\mathcal{H}_\ell^{(1)}(M)$ as *principal* block-submatrices guarantees that $\widetilde{X}_{ij}$ and $\widetilde{Y}_{ij}$ are block-symmetric. Similarly, defining the block-symmetric matrices

$$\widehat{X}_{ij} = \begin{array}{c c} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ j \end{array} & \begin{bmatrix} & & \\ & \boxed{-\mathcal{H}_\ell^{(-1)}(P)} & \\ & & \end{bmatrix} \end{array}, \quad \widehat{Y}_{ij} = \begin{array}{c c} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ j \end{array} & \begin{bmatrix} & & \\ & \boxed{\mathcal{H}_\ell^{(0)}(P)} & \\ & & \end{bmatrix} \end{array}, \tag{3.2.6}$$

we have

$$\widehat{X}_{ij} \boxplus\!\!\rightarrow \widehat{Y}_{ij} = E_{i,j+1} \otimes P \qquad (i \le j). \tag{3.2.7}$$

Thus sums of these principally embedded versions of (3.2.3) and (3.2.4) can produce an arbitrary block $k \times (k+1)$ matrix Z as the column-shifted sum of block-symmetric X and Y. □

Lemma 3.2.4. *Suppose X and Y are both block-symmetric block $k \times k$ matrices with $n \times n$ blocks. Then $X \boxplus\!\!\rightarrow Y = 0 \iff X = Y = 0$.*

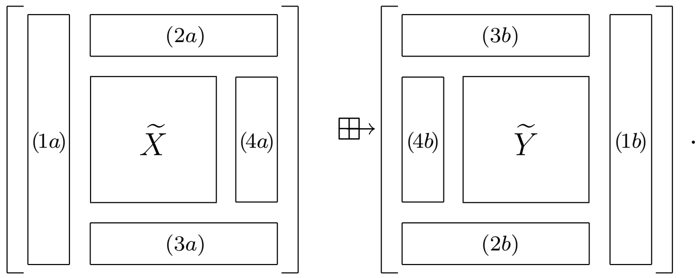
Proof. The proof is by induction on k. We focus on the nontrivial direction ($\Rightarrow$). There are two base cases to be checked, $k = 1$ and $k = 2$. The $k = 1$ case is immediate. Because X and Y are block-symmetric, for $k = 2$ we have

$$X \boxplus\!\!\rightarrow Y = \begin{bmatrix} X_{11} & X_{12} & 0 \\ X_{12} & X_{22} & 0 \end{bmatrix} + \begin{bmatrix} 0 & Y_{11} & Y_{12} \\ 0 & Y_{12} & Y_{22} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} + Y_{11} & Y_{12} \\ X_{12} & X_{22} + Y_{12} & Y_{22} \end{bmatrix}.$$

Then $X \boxplus\!\!\rightarrow Y = 0$ clearly implies that $X = Y = 0$.

Now consider $k > 2$ and X and Y with their blocks "around the edges" grouped together as indicated in the diagram:

$X \boxplus\!\!\rightarrow Y =$



The only contribution to the first block column of $X \boxplus\!\!\rightarrow Y$ comes from $(1a)$, and the only contribution to the last block column of $X \boxplus\!\!\rightarrow Y$ comes from $(1b)$. Thus $X \boxplus\!\!\rightarrow Y = 0$ implies $(1a)$ and $(1b)$ are all zeros. (Note that this would be true for general X and Y.) The block-symmetry of X and Y now implies that the blocks in $(2a)$ and $(2b)$ are zero. The blocks of $(2a)$ interact in the shifted sum with those in $(3b)$; the $(2a)$ blocks being zero imply that all the $(3b)$ blocks are zero. Similarly the $(2b)$ blocks all zero imply that all the $(3a)$ blocks are zero. Finally, the block-symmetry of X and Y can be invoked once again to see that all the $(4a)$ and $(4b)$ blocks are zero. At this point we have that $X \boxplus\!\!\rightarrow Y = 0$ implies

$$X \boxplus\!\!\rightarrow Y = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \widetilde{X} & \vdots \\ 0 & \dots & 0 \end{bmatrix} \boxplus\!\!\rightarrow \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \widetilde{Y} & \vdots \\ 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \widetilde{X} \boxplus\!\!\rightarrow \widetilde{Y} & \vdots \\ 0 & \dots & 0 \end{bmatrix} = 0.$$

Since $\widetilde{X} \boxplus\!\!\rightarrow \widetilde{Y} = 0$, the induction hypothesis implies $\widetilde{X} = \widetilde{Y} = 0$, and consequently that $X = Y = 0$. $\square$

The results of this section are concisely summarized by the following corollary of Lemmas 3.2.3 and 3.2.4.

Corollary 3.2.5. *Let Z be an arbitrary block $k \times (k+1)$ matrix with $n \times n$ blocks. Then there exist unique block-symmetric block $k \times k$ matrices X and Y with $n \times n$ blocks such that $X \boxplus\!\!\rightarrow Y = Z$.*

3.3 Block-symmetric Pencils in $\mathbb{L}_1(P)$

We are now in a position to characterize the subspace $\mathbb{B}(P)$ of all block-symmetric pencils in $\mathbb{L}_1(P)$, and to give explicit formulas for a useful basis of this subspace.

3.3.1 The Subspace $\mathbb{B}(P)$

Using the results of Section 3.2.1 we now characterize the subspace $\mathbb{B}(P) \subseteq \mathbb{L}_1(P)$. Later in Section 5.3 we will see that almost all of these pencils are indeed linearizations for P.

Theorem 3.3.1 (Characterization of $\mathbb{B}(P)$).
For any matrix polynomial $P(\lambda)$ of degree k, $\dim \mathbb{B}(P) = k$, and for each vector $v \in \mathbb{F}^k$ there is a uniquely determined block-symmetric pencil in $\mathbb{B}(P)$.

Proof. The theorem is proved if we can show that the restriction to $\mathbb{B}(P)$ of the multiplication map (2.1.9), that is

$$\begin{array}{rcl} \mathbb{B}(P) & \xrightarrow{\mathcal{M}} & \mathcal{V}_P := \{\, v \otimes P(\lambda) : v \in \mathbb{F}^k \,\} \\ L(\lambda) & \longmapsto & L(\lambda)\,(\Lambda \otimes I_n)\ , \end{array} \tag{3.3.1}$$

is a linear isomorphism.

First, recall from Lemma 2.1.4 that for any pencil $\lambda X + Y \in \mathbb{L}_1(P)$,

$$(\lambda X + Y)(\Lambda \otimes I_n) = v \otimes P(\lambda) \quad \Longleftrightarrow \quad X \boxplus\!\!\rightarrow Y = v \otimes [\, A_k \ A_{k-1} \ \dots \ A_0 \,]. \tag{3.3.2}$$

Thus $\lambda X + Y$ is in $\ker \mathcal{M}$ iff $X \boxplus\!\!\rightarrow Y = 0$. But X and Y are block-symmetric, so by Lemma 3.2.4 we see that $\ker \mathcal{M} = \{0\}$, and hence $\mathcal{M}$ is 1-1.

To see that $\mathcal{M}$ is onto, let $v \otimes P(\lambda)$ with $v \in \mathbb{F}^k$ be an arbitrary element of $\mathcal{V}_P$. With $Z = v \otimes [A_k \ A_{k-1} \ \dots \ A_0]$, the construction of Lemma 3.2.3 shows that there exist block-symmetric X and Y such that $X \boxplus\!\!\rightarrow Y = v \otimes [A_k \ A_{k-1} \ \dots \ A_0]$. Then by (3.3.2) we have $\mathcal{M}(\lambda X + Y) = v \otimes P(\lambda)$, showing that $\mathcal{M}$ is onto. ∎

3.3.2 The "Standard Basis" for $\mathbb{B}(P)$

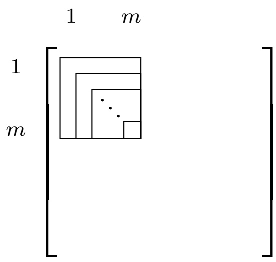
The isomorphism established in the proof of Theorem 3.3.1 immediately suggests the possibility that the basis for $\mathbb{B}(P)$ corresponding (via the map $\mathcal{M}$ in (3.3.1)) to the standard basis $\{e_1, \dots, e_k\}$ for $\mathbb{F}^k$ may be especially simple and useful. In this section we derive a general formula for these "standard basis pencils" in $\mathbb{B}(P)$ as a corollary of the shifted sum construction used in the proof of Lemma 3.2.3.

In light of Lemma 2.1.4, then, our goal is to construct for each $1 \le m \le k$ a block-symmetric pencil $\lambda X_m + Y_m$ such that

$$X_m \boxplus\!\!\rightarrow Y_m \ = \ e_m \otimes [\, A_k \ A_{k-1} \ \dots \ A_0 \,]. \tag{3.3.3}$$

This is most easily done in two separate steps. First we show how to achieve the initial m block-columns in the desired shifted sum, i.e., how to get $e_m \otimes [\, A_k \ \dots \ A_{k-m+1} \ 0 \ \dots \ 0\,]$. Then the final $k-m+1$ block-columns $e_m \otimes [\, 0 \ \dots \ 0 \ A_{k-m} \ \dots \ A_1 \ A_0 \,]$ are produced by a related but slightly different construction. In each of these constructions we use the following notation for principal block submatrices, adapted from [40]: for a block $k \times k$ matrix X and index set $\alpha \subseteq \{1, 2, \dots, k\}$, $X(\alpha)$ will denote the principal block submatrix lying in the block rows and block columns with indices in α.

To get the first m block-columns in the desired shifted sum we repeatedly use the construction in (3.2.5) to build block $k \times k$ matrices $\widehat{X}_m$ and $\widehat{Y}_m$, embedding once in each of the principal block submatrices $\widehat{X}_m(\alpha_i)$ and $\widehat{Y}_m(\alpha_i)$ for the index sets $\alpha_i = \{i, i+1, \ldots, m\}$, $i = 1\colon m$, pictured in the following diagram.



(3.3.4)

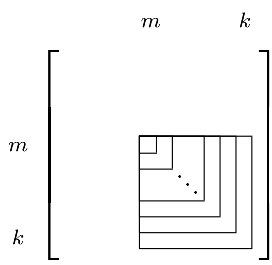
Accumulating these embedded submatrices, we obtain

$$\widehat{X}_m = \begin{array}{c} \begin{array}{ccc} & m & \end{array} \\ \left[\begin{array}{cccc|c} & & & A_k & \\ & & \iddots & A_{k-1} & 0 \\ & \iddots & \iddots & \vdots & \\ \iddots & \iddots & & \vdots & \\ A_k & A_{k-1} & \cdots\cdots & A_{k-m+1} & \\ \hline & & 0 & & 0 \end{array}\right] \end{array} m \quad ,$$

$$\widehat{Y}_m = - \begin{array}{c} \begin{array}{ccc} & & m \end{array} \\ \left[\begin{array}{ccccc|c} & & & A_k & 0 & \\ & & \iddots & A_{k-1} & 0 & \\ & \iddots & \iddots & \vdots & \vdots & 0 \\ A_k & A_{k-1} & \ldots & A_{k-m+2} & 0 & \\ 0 & \ldots & \ldots & 0 & 0 & \\ \hline & & 0 & & & 0 \end{array}\right] \end{array} m \quad ,$$

with the property that $\widehat{X}_m \boxplus\!\!\rightarrow \widehat{Y}_m = e_m \otimes [\, A_k \ \ldots \ A_{k-m+1} \ \ 0 \ \ldots \ 0 \,]$.

To obtain the last $k-m+1$ columns we use the construction outlined in (3.2.6) and (3.2.7) $k-m+1$ times to build block $k \times k$ matrices $\widetilde{X}_m$ and $\widetilde{Y}_m$, embedding once in each of the principal block submatrices $\widetilde{X}_m(\beta_j)$ and $\widetilde{Y}_m(\beta_j)$ for the index sets $\beta_j = \{m, m+1, \ldots, j\}$ where $j = m\colon k$, as in diagram (3.3.5).



(3.3.5)

This yields

$$\widetilde{X}_m = - \begin{array}{c} \begin{array}{cc} \hspace{2em} m & \hspace{6em} k \end{array} \\ \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \begin{matrix} 0 & 0 & \cdots\cdots & 0 \\ 0 & A_{k-m-1} & \cdots & A_1 \; A_0 \\ \vdots & \vdots & \iddots & \iddots \\ \vdots & A_1 & \iddots & \\ 0 & A_0 & & \end{matrix} \end{array}\right] \end{array} \begin{array}{c} \\ m \\ \\ k \end{array},$$

$$\widetilde{Y}_m = \begin{array}{c} \begin{array}{cc} \hspace{2em} m & \hspace{6em} k \end{array} \\ \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \begin{matrix} A_{k-m} & \cdots\cdots & A_1 \; A_0 \\ \vdots & \iddots & A_0 \\ \vdots & \iddots \; \iddots & \\ A_1 & \iddots & \\ A_0 & & \end{matrix} \end{array}\right] \end{array} \begin{array}{c} \\ m \\ \\ k \end{array},$$

satisfying $\widetilde{X}_m \boxplus\!\!\rightarrow \widetilde{Y}_m = e_m \otimes [\, 0 \ \ldots \ 0 \ A_{k-m} \ \ldots \ A_1 \ A_0 \,]$. With $X_m := \widehat{X}_m + \widetilde{X}_m$ and $Y_m := \widehat{Y}_m + \widetilde{Y}_m$ we have $X_m \boxplus\!\!\rightarrow Y_m = e_m \otimes [\, A_k \ A_{k-1} \ \ldots \ A_1 \ A_0 \,]$, so $\lambda X_m + Y_m$ is the mth standard basis pencil for $\mathbb{B}(P)$.

A more concise way to express the mth standard basis pencil uses the following block-Hankel matrices. Let $\mathcal{L}_j(P(\lambda))$ denote the lower block-anti-triangular, block-Hankel, block $j \times j$ matrix

$$\mathcal{L}_j(P(\lambda)) := \begin{bmatrix} & & A_k \\ & \iddots & A_{k-1} \\ \iddots & \iddots & \vdots \\ A_k \ A_{k-1} & \ldots & A_{k-j+1} \end{bmatrix} \tag{3.3.6}$$

formed from the first j matrix coefficients $A_k, A_{k-1}, \ldots, A_{k-j+1}$ of $P(\lambda)$. Similarly, let $\mathcal{U}_j(P(\lambda))$ denote the upper block-anti-triangular, block-Hankel, block $j \times j$ matrix

$$\mathcal{U}_j(P(\lambda)) := \begin{bmatrix} A_{j-1} & \ldots & A_1 \ A_0 \\ \vdots & \iddots & \iddots \\ A_1 & \iddots & \\ A_0 & & \end{bmatrix} \tag{3.3.7}$$

formed from the last j matrix coefficients $A_{j-1}, A_{j-2}, \ldots, A_1, A_0$ of $P(\lambda)$. Then the block-symmetric matrices X_m and Y_m in the mth standard basis pencil ($m = 1\colon k$) can be neatly expressed as a direct sum of block-Hankel matrices:

$$X_m = X_m(P(\lambda)) = \begin{bmatrix} \mathcal{L}_m(P(\lambda)) & 0 \\ 0 & -\mathcal{U}_{k-m}(P(\lambda)) \end{bmatrix}, \tag{3.3.8a}$$

$$Y_m = Y_m(P(\lambda)) = \begin{bmatrix} -\mathcal{L}_{m-1}(P(\lambda)) & 0 \\ 0 & \mathcal{U}_{k-m+1}(P(\lambda)) \end{bmatrix}. \tag{3.3.8b}$$

Table 3.3.1: "Standard basis pencils" in $\mathbb{B}(Q)$ for quadratic $Q(\lambda) = \lambda^2 A + \lambda B + C$.

v	$L(\lambda) \in \mathbb{B}(Q)$
$\begin{bmatrix}1\\0\end{bmatrix}$	$\lambda \left[\begin{array}{c\|c} A & 0 \\ \hline 0 & -C \end{array}\right] + \begin{bmatrix} B & C \\ C & 0 \end{bmatrix}$
$\begin{bmatrix}0\\1\end{bmatrix}$	$\lambda \begin{bmatrix} 0 & A \\ A & B \end{bmatrix} + \left[\begin{array}{c\|c} -A & 0 \\ \hline 0 & C \end{array}\right]$

($\mathcal{L}_j$ and $\mathcal{U}_j$ are taken to be void when $j = 0$.) From (3.3.8) it now becomes obvious that the coefficient matrices in successive standard basis pencils are closely related:

$$Y_m(P(\lambda)) \;=\; -X_{m-1}(P(\lambda)), \qquad m = 1\colon k. \tag{3.3.9}$$

Thus we have the following explicit formula for the standard basis pencils in $\mathbb{B}(P)$.

Theorem 3.3.2 (Standard Basis for $\mathbb{B}(P)$).
Let $P(\lambda)$ be a matrix polynomial of degree k. Then for $m = 1\colon k$ the block-symmetric pencil in $\mathbb{B}(P)$ with ansatz vector e_m is $\lambda X_m - X_{m-1}$, where X_m is given by (3.3.8a).

The standard basis pencils in $\mathbb{B}(P)$ for general polynomials of degree 2 and 3 are listed in Tables 3.3.1 and 3.3.2, where the partitioning from (3.3.8) is shown in each case. As an immediate consequence we have, for the important case of quadratic polynomials $Q(\lambda) = \lambda^2 A + \lambda B + C$, the following description of all block-symmetric pencils in $\mathbb{L}_1(Q)$,

$$\mathbb{B}(Q) = \left\{ L(\lambda) = \lambda \begin{bmatrix} v_1 A & v_2 A \\ v_2 A & v_2 B - v_1 C \end{bmatrix} + \begin{bmatrix} v_1 B - v_2 A & v_1 C \\ v_1 C & v_2 C \end{bmatrix} \;:\; v \in \mathbb{C}^2 \right\}.$$

v	$L(\lambda) \in \mathbb{B}(P)$
$\begin{bmatrix}1\\0\\0\end{bmatrix}$	$\lambda \left[\begin{array}{c\|cc} A & 0 & 0 \\ \hline 0 & -C & -D \\ 0 & -D & 0 \end{array}\right] + \begin{bmatrix} B & C & D \\ C & D & 0 \\ D & 0 & 0 \end{bmatrix}$
$\begin{bmatrix}0\\1\\0\end{bmatrix}$	$\lambda \left[\begin{array}{cc\|c} 0 & A & 0 \\ A & B & 0 \\ \hline 0 & 0 & -D \end{array}\right] + \left[\begin{array}{c\|cc} -A & 0 & 0 \\ \hline 0 & C & D \\ 0 & D & 0 \end{array}\right]$
$\begin{bmatrix}0\\0\\1\end{bmatrix}$	$\lambda \begin{bmatrix} 0 & 0 & A \\ 0 & A & B \\ A & B & C \end{bmatrix} + \left[\begin{array}{cc\|c} 0 & -A & 0 \\ -A & -B & 0 \\ \hline 0 & 0 & D \end{array}\right]$

Table 3.3.2: "Standard basis pencils" in $\mathbb{B}(P)$ for cubic $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$.

3.4 Double Ansatz Pencils for General P

With a complete characterization of $\mathbb{B}(P)$ in hand, an efficient development of the space $\mathbb{DL}(P)$ of double ansatz pencils can now be given. By definition, a pencil $L(\lambda) \in \mathbb{DL}(P)$ simultaneously satisfies both

$$\text{a "right ansatz"} \qquad L(\lambda) \cdot (\Lambda \otimes I) = v \otimes P(\lambda) \tag{3.4.1}$$

$$\text{and a "left ansatz"} \qquad (\Lambda^T \otimes I) \cdot L(\lambda) = w^T \otimes P(\lambda) \tag{3.4.2}$$

for some vectors $v, w \in \mathbb{F}^k$. But can every vector pair (v, w) actually be realized as the right/left ansatz vectors of some $\mathbb{DL}(P)$-pencil? From Theorem 3.1.6, Theorem 3.3.1, and the containment $\mathbb{B}(P) \subseteq \mathbb{DL}(P)$ from Corollary 3.2.2 we know that any pair (v, w) with $w = v$ is achieved by a block-symmetric pencil in $\mathbb{DL}(P)$. But are any other pairs attained? Indeed, are there ever any pencils in $\mathbb{DL}(P)$ that are *not* block-symmetric?

In this section we show that the answer to all these questions is NO ; *every* pencil in $\mathbb{DL}(P)$ is block-symmetric, i.e., the containment $\mathbb{B}(P) \subseteq \mathbb{DL}(P)$ is actually an equality. We begin by considering the special case of $\mathbb{DL}(P)$-pencils with right ansatz vector $v = 0$, showing that in this case $w = 0$ is forced and the pencil is unique.

Lemma 3.4.1. *Suppose $L(\lambda) = \lambda X + Y \in \mathbb{DL}(P)$ has right ansatz vector v and left ansatz vector w. Then $v = 0$ implies that w must also be 0, and that $X = Y = 0$.*

Proof. We first show that the ℓth block-column of X and the ℓth coordinate of w is zero for $\ell = 1\colon k$ by an induction on ℓ.

Suppose $\ell = 1$. From Lemma 2.1.4 we know that $X \boxplus\!\!\rightarrow Y = v \otimes [A_k \; A_{k-1} \; \dots \; A_0]$. Since $v = 0$ we have $X \boxplus\!\!\rightarrow Y = 0$, and hence the first block-column of X is zero. But from Theorem 3.1.6, $L(\lambda)$ being in $\mathbb{L}_2(P)$ with left ansatz vector w implies that $L(\lambda)^{\mathcal{B}} \in \mathbb{L}_1(P)$ with right ansatz vector w, which can be written in terms of the shifted sum as

$$X^{\mathcal{B}} \boxplus\!\!\rightarrow Y^{\mathcal{B}} = w \otimes [A_k \; A_{k-1} \; \dots \; A_0]. \tag{3.4.3}$$

The $(1,1)$-block of the right-hand side of (3.4.3) is $w_1 A_k$, while on the left-hand side the $(1,1)$-block of $X^{\mathcal{B}} \boxplus\!\!\rightarrow Y^{\mathcal{B}}$ is the same as the $(1,1)$-block of X. Hence $w_1 A_k = 0$. But the leading coefficient A_k of $P(\lambda)$ is nonzero by assumption, so $w_1 = 0$.

Now suppose that the ℓth block-column of X is zero and that $w_\ell = 0$. Then by (3.4.3) the ℓth block-row of $X^{\mathcal{B}} \boxplus\!\!\rightarrow Y^{\mathcal{B}}$ is zero, which together with the ℓth block-row of $X^{\mathcal{B}}$ being zero implies that the ℓth block-row of $Y^{\mathcal{B}}$ is zero or equivalently, the ℓth block-column of Y is zero. Combining this with $X \boxplus\!\!\rightarrow Y = 0$ implies that the $(\ell+1)$th block-column of X is zero. Now equating the $(\ell+1, 1)$-blocks of both sides of (3.4.3) gives $w_{\ell+1} A_k = 0$, and hence $w_{\ell+1} = 0$. This concludes the induction, and shows that $X = 0$ and $w = 0$.

Finally, $X = 0$ and $X \boxplus\!\!\rightarrow Y = 0$ implies $Y = 0$, completing the proof. ∎

We can now give a precise description of all right/left ansatz vector pairs (v, w) that can be realized by some $\mathbb{DL}(P)$-pencil. Consequently $\mathbb{DL}(P)$ is characterized as being identical to $\mathbb{B}(P)$, and the main goal of this chapter is achieved.

Theorem 3.4.2 (Characterization of $\mathbb{DL}(P)$).
For a matrix polynomial $P(\lambda)$ of degree k, suppose $L(\lambda) \in \mathbb{DL}(P)$ with right ansatz

vector v and left ansatz vector w. Then $v = w$ and $L(\lambda) \in \mathbb{B}(P)$. Thus $\mathbb{DL}(P) = \mathbb{B}(P)$, $\dim \mathbb{DL}(P) = k$, and for each $v \in \mathbb{F}^k$ there is a uniquely determined pencil in $\mathbb{DL}(P)$.

Proof. Let $\mathcal{L}(\lambda) \in \mathbb{B}(P)$ be the unique block-symmetric pencil from Theorem 3.3.1 with v as its right ansatz vector. From Theorem 3.1.6 we know that $\mathcal{L}(\lambda)^{\mathcal{B}} = \mathcal{L}(\lambda)$ is in $\mathbb{L}_2(P)$ with left ansatz vector v, and so $\mathcal{L}(\lambda) \in \mathbb{DL}(P)$ with v as both its right and left ansatz vector. Thus the pencil $\widetilde{L}(\lambda) := L(\lambda) - \mathcal{L}(\lambda)$ is in $\mathbb{DL}(P)$ with right ansatz vector 0 and left ansatz vector $w - v$. Lemma 3.4.1 then implies that $v = w$ and $\widetilde{L}(\lambda) = \lambda \cdot 0 + 0$. Thus $L(\lambda) \equiv \mathcal{L}(\lambda) \in \mathbb{B}(P)$, so $\mathbb{DL}(P) \subseteq \mathbb{B}(P)$. In view of Corollary 3.2.2 we can conclude that $\mathbb{DL}(P) = \mathbb{B}(P)$. The rest of the theorem follows immediately from the characterization of $\mathbb{B}(P)$ in Theorem 3.3.1. ∎

The equality $\mathbb{DL}(P) = \mathbb{B}(P)$ can be thought of as saying that the pencils in $\mathbb{DL}(P)$ are *doubly* structured: they have block-symmetry as well as the eigenvector recovery properties that were the original motivation for their definition. The equality also means that the basis of $\mathbb{B}(P)$ developed in section 3.3.2 is also, of course, a "standard basis" for $\mathbb{DL}(P)$.

Note that because of the equality of right and left ansatz vectors v and w for any pencil in $\mathbb{DL}(P)$, we can (and will) from now on refer without ambiguity to *the ansatz vector* v of $L(\lambda)$, whenever $L(\lambda) \in \mathbb{DL}(P)$.

3.5 Some Other Constructions of Block-symmetric Linearizations

Several other methods for constructing block-symmetric linearizations of matrix polynomials have appeared previously in the literature.

Antoniou and Vologiannidis [3] have recently found new companion-like linearizations for general matrix polynomials P by generalizing Fiedler's results [25] on a factorization of the companion matrix of a scalar polynomial and certain of its permutations. From this finite family of $\frac{1}{6}(2 + \deg P)!$ pencils, all of which are linearizations, they identify one distinguished pencil that is Hermitian whenever P is Hermitian. But this example has structure even for general P: it is block-symmetric. Indeed, it provides a simple example of a block-symmetric linearization for $P(\lambda)$ that is not in $\mathbb{B}(P)$. In the case of a cubic polynomial $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$, the pencil is

$$L(\lambda) = \lambda \begin{bmatrix} A & 0 & 0 \\ 0 & 0 & I \\ 0 & I & C \end{bmatrix} + \begin{bmatrix} B & -I & 0 \\ -I & 0 & 0 \\ 0 & 0 & D \end{bmatrix}. \tag{3.5.1}$$

Using the column-shifted sum it easy to see that $L(\lambda)$ is not in $\mathbb{L}_1(P)$, and hence not in $\mathbb{B}(P)$.

Contrasting with the "permuted factors" approach of [3],[25] and the additive construction used in this thesis, is a third "multiplicative" method for generating block-symmetric linearizations described by Lancaster in [46], [47]. In [46] only scalar polynomials $p(\lambda) = a_k\lambda^k + \cdots + a_1\lambda + a_0$ are considered; the starting point is the companion matrix of $p(\lambda)$,

$$C = \begin{bmatrix} -a_k^{-1} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \begin{bmatrix} a_{k-1} & a_{k-2} & \dots & a_0 \\ 1 & 0 & \dots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & 1 & 0 \end{bmatrix} \tag{3.5.2}$$

and the associated pencil $\lambda I - C$. Lancaster's strategy is to seek a nonsingular symmetric matrix B such that BC is symmetric, thus providing a symmetric linearization $B(\lambda I - C) = \lambda B - BC$ for $p(\lambda)$. That such a B can always be found follows from a standard result in matrix theory [40, Cor. 4.4.11]. Lancaster shows further that B and BC symmetric implies BC^j is symmetric for all $j \geq 1$; thus $BC^{j-1}(\lambda I - C) = \lambda BC^{j-1} - BC^j$ is a symmetric pencil for any $j \geq 1$, and for $j \geq 2$ it is a linearization of $p(\lambda)$ if $a_0 \neq 0$. This strategy is realized in [46] with the particular choice of symmetric (Hankel) matrix

$$B = \begin{bmatrix} & & & a_k \\ & & \cdot^{\cdot^{\cdot}} & a_{k-1} \\ & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots \\ a_k & a_{k-1} & \dots & a_1 \end{bmatrix}, \tag{3.5.3}$$

which is nonsingular since $a_k \neq 0$. It is observed that this particular B gives a sequence of symmetric pencils

$$\lambda BC^{j-1} - BC^j, \quad j = 1, \infty \tag{3.5.4}$$

with an especially simple form for $1 \leq j \leq k$, though apparently with a much more complicated form for $j > k$.

It is easy to see that these symmetric pencils, constructed for scalar polynomials $p(\lambda)$, can be immediately extended to block-symmetric pencils for general matrix polynomials $P(\lambda)$ simply by formally replacing the scalar coefficients of $p(\lambda)$ in $B, BC, BC^2, \dots$ by the matrix coefficients of $P(\lambda)$. This has been done in [47, Sect. 4.2] and [29]. Garvey et al. [29] go even further with these block-symmetric pencils, using them as a foundation for defining a new class of isospectral transformations on matrix polynomials.

Since Lancaster's construction of pencils is so different from ours there is no a priori reason to expect any connection between his pencils and the pencils in $\mathbb{DL}(P)$. The next result shows, rather surprisingly, that the first k pencils in the sequence (3.5.4) generate $\mathbb{DL}(P)$.

Theorem 3.5.1. *Let $P(\lambda)$ be any matrix polynomial of degree k. Then for $m = 1\colon k$ the pencil $\lambda BC^{k-m} - BC^{k-m+1}$ from the sequence* (3.5.4)*, with B and C defined by the block matrix analogs of* (3.5.2) *and* (3.5.3)*, is identical to $\lambda X_m - X_{m-1}$, the mth standard basis pencil for $\mathbb{DL}(P)$.*

Proof. We have to show that $X_m = BC^{k-m}$ for $m = 0\colon k$, where X_m is given by (3.3.8a). For notational simplicity we will carry out the proof for a scalar polynomial; the same proof applies to a matrix polynomial with only minor changes in notation. The $m = k$ case, $X_k(p(\lambda)) = \mathcal{L}_k(p(\lambda)) = B$, is immediate from equations

(3.3.6), (3.3.8), and (3.5.3). The rest follow inductively (downward) from the relation $X_{m-1}(p(\lambda)) = X_m(p(\lambda)) \cdot C$, which we now proceed to show holds for $m = 1\colon k$.

To see that $X_m C = X_{m-1}$, or equivalently that

$$\begin{bmatrix} \mathcal{L}_m(p(\lambda)) & 0 \\ 0 & -\mathcal{U}_{k-m}(p(\lambda)) \end{bmatrix} C = \begin{bmatrix} \mathcal{L}_{m-1}(p(\lambda)) & 0 \\ 0 & -\mathcal{U}_{k-m+1}(p(\lambda)) \end{bmatrix}$$

holds for $m = 1\colon k$, it will be convenient to rewrite the companion matrix (3.5.2) in the form

$$C = N_k^T - a_k^{-1} \begin{bmatrix} a_{k-1} & a_{k-2} & \dots & a_0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = N_k^T - a_k^{-1} e_1 \begin{bmatrix} a_{k-1} & a_{k-2} & \dots & a_0 \end{bmatrix},$$

where N_k is defined in (3.2.2). Then

$$\begin{aligned} X_m(p(\lambda))C &= X_m(p(\lambda))N_k^T - a_k^{-1} X_m(p(\lambda))\, e_1 \begin{bmatrix} a_{k-1} & a_{k-2} & \dots & a_0 \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{L}_m(p(\lambda)) & 0 \\ 0 & -\mathcal{U}_{k-m}(p(\lambda)) \end{bmatrix} N_k^T - e_m \begin{bmatrix} a_{k-1} & a_{k-2} & \dots & a_0 \end{bmatrix}. \end{aligned}$$

In the first term, postmultiplication by N_k^T has the effect of shifting the columns to the left by one (and losing the first column), thus giving

$$\begin{aligned} X_m(p(\lambda))C &= \\ & \begin{bmatrix} \mathcal{L}_{m-1}(p(\lambda)) & 0 & 0 \\ a_{k-1} \dots a_{k-m+1} & 0 & 0 \\ 0 & -\mathcal{U}_{k-m}(p(\lambda)) & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ a_{k-1} \dots a_{k-m+1} & a_{k-m} \dots a_1 & a_0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \left[\begin{array}{c|cc} \mathcal{L}_{m-1}(p(\lambda)) & 0 & 0 \\ \hline 0 & -a_{k-m} \dots - a_1 & -a_0 \\ 0 & -\mathcal{U}_{k-m}(p(\lambda)) & 0 \end{array}\right] \\ &= \begin{bmatrix} \mathcal{L}_{m-1}(p(\lambda)) & 0 \\ 0 & -\mathcal{U}_{k-m+1}(p(\lambda)) \end{bmatrix} = X_{m-1}(p(\lambda)). \end{aligned}$$

This completes the inductive step of the proof. $\square$

Chapter 4

The Genericity of Linearizations in $\mathbb{DL}(P)$

In this chapter we reconsider the symbolic "linearization condition" discussed in Section 2.2.2. As illustrated by Example 2.2.6, the intrinsic meaning of this condition can sometimes be rather obscure. However, we will see that for pencils in $\mathbb{DL}(P)$ this condition can always be expressed in a way that makes its meaning transparent. Indeed, one of the most striking properties of the space $\mathbb{DL}(P)$ is that the linearization condition for each $\mathbb{DL}(P)$-pencil can be directly linked to its ansatz vector v, as will be seen in the following sections. An important consequence of this link is that "almost every" pencil in $\mathbb{DL}(P)$ is a linearization for P.

4.1 Some Suggestive Examples

Let us begin by considering some concrete examples. Tables 4.1.1 and 4.1.2 display a sampling of double ansatz pencils for general quadratic and cubic matrix polynomials, together with their corresponding linearization conditions. Looking at just the standard basis pencils in these tables, it is not so easy to discern any clear pattern in the linearization conditions. However, the last two entries of Table 4.1.1 present a substantial clue; for quadratic polynomials the linearization condition can always be interpreted as excluding one particular value, $\lambda = -\beta/\alpha$, as an eigenvalue of P. More precisely, the $\mathbb{DL}(P)$-pencil with ansatz vector $v = \left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]$ is a linearization for P if and only if $\lambda = -\beta/\alpha$ is *not* an eigenvalue of P. This is a clear indication of a strong connection between the linearization condition of a $\mathbb{DL}(P)$-pencil and its ansatz vector. But is there still such a strong connection for higher degree polynomials, and if so what is the nature of that link?

At first glance the examples of $\mathbb{DL}(P)$-pencils in Table 4.1.2 seem to shed no light on this question. The linearization conditions, especially for the last two examples, do not appear to have any clear connection at all with their corresponding ansatz vectors. However, let's examine the fourth example in more detail. For $L(\lambda) \in \mathbb{DL}(P)$ with ansatz vector $v = \left[\begin{smallmatrix}1\\ 0\\ -1\end{smallmatrix}\right]$, one easily finds using the procedure in Section 2.2.2 that

$$\det \begin{bmatrix} A+C & B+D \\ B+D & A+C \end{bmatrix} \neq 0 \tag{4.1.1}$$

Table 4.1.1: Some pencils in $\mathbb{DL}(P)$ for the general quadratic $P(\lambda) = \lambda^2 A + \lambda B + C$. Linearization condition found using procedure in Section 2.2.2.

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

Table 4.1.2: Some pencils in $\mathbb{DL}(P)$ for the general cubic $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$. Linearization condition found using procedure in Section 2.2.2.

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

is the linearization condition for $L(\lambda)$. Now it is not immediately clear what the meaning of this condition is, whether it has any connection to the ansatz vector v, or even if it has any intrinsic meaning at all. However, the identity

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

shows that condition (4.1.1) is equivalent to saying that neither -1 nor $+1$ is an eigenvalue of the matrix polynomial $P(\lambda)$. Thus in this example we can again interpret the linearization condition from Section 2.2.2 as an "eigenvalue exclusion" condition.

But why do these particular eigenvalues need to be excluded? And what role, if any, does the ansatz vector $v = \left[\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right]$ play here? Observe that if we interpret the components of v as the coefficients of a scalar polynomial then we obtain $x^2 - 1$, whose roots are exactly the eigenvalues that have to be excluded in order to guarantee that $L(\lambda)$ is a linearization for $P(\lambda)$. Similarly in the quadratic case, the excluded value $\lambda = -\beta/\alpha$ is the root of the scalar polynomial $\alpha x + \beta$ obtained from the components of the ansatz vector $v = \left[\begin{smallmatrix} \alpha \\ \beta \end{smallmatrix}\right]$. One of the goals of this chapter is to show that these are not merely coincidences, but rather instances of a general phenomenon described by the "eigenvalue exclusion theorem".

4.2 Determinant of $\mathbb{DL}(P)$-pencils

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

Definition 4.2.1 (v-polynomial).
To a vector $v = [\, v_1, v_2, \ldots, v_k\,]^T \in \mathbb{F}^k$ associate the scalar polynomial

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

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

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

Definition 4.2.2 (Horner shifts).
For any polynomial $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ and $0 \le \ell \le n$, the "degree ℓ Horner shift of $p(x)$" is $p_\ell(x) := a_n x^\ell + a_{n-1} x^{\ell-1} + \cdots + a_{n-\ell+1} x + a_{n-\ell}$.

Remark 4.2.3. *The polynomials in Definition* 4.2.2 *satisfy the recurrence relation*

$$\begin{aligned} p_0(x) &= a_n, \\ p_{\ell+1}(x) &= x p_\ell(x) + a_{n-\ell-1} \quad \textit{for } 0 \le \ell \le n-1, \\ p_n(x) &= p(x)\,, \end{aligned}$$

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

We have seen in Theorem 3.4.2 that $L(\lambda) \in \mathbb{DL}(P)$ is uniquely determined by the vector v and the polynomial P, so it is not surprising that one can also specify the columns of $L(\lambda)$ in terms of this data. This is done in the next lemma, where a block-column-wise description of $L(\lambda)$ is given. In this description we make extensive use of the standard $k \times k$ nilpotent Jordan block N from (3.2.2) in the matrix $N \otimes I$, employed here as a block-shift operator.

Lemma 4.2.4 (Block-column structure of pencils in $\mathbb{DL}(P)$).
Suppose that $L(\lambda) = \lambda X + Y$ *is in* $\mathbb{DL}(P)$ *with ansatz vector* v*. Partition* X *and* Y *into block-columns*

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

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

$$Y_\ell = (N \otimes I)(Y_{\ell-1} - v \otimes A_{k-\ell+1}) + v_\ell \begin{bmatrix} A_{k-1} \\ \vdots \\ A_0 \end{bmatrix} \qquad 1 \le \ell \le k-1, \tag{4.2.1}$$

$$Y_k = v \otimes A_0. \tag{4.2.2}$$

The block-columns of X *are then determined by the relation*

$$X_\ell = -Y_{\ell-1} + v \otimes A_{k-\ell+1} \quad \textit{for} \quad 1 \le \ell \le k\,, \tag{4.2.3}$$

and the pencil $L(\lambda)$ *has the block-column-wise description*

$$L(\lambda) = \left[\begin{array}{l|l|c|l|l} Y_1 & Y_2 - \lambda Y_1 & \cdots & Y_{k-1} - \lambda Y_{k-2} & v \otimes A_0 - \lambda Y_{k-1} \\ \quad + \lambda v \otimes A_k & \quad + \lambda v \otimes A_{k-1} & & \quad + \lambda v \otimes A_2 & \quad + \lambda v \otimes A_1 \end{array}\right]. \tag{4.2.4}$$

The topmost $(n \times n)$ *blocks of the block-columns* Y_ℓ *are given by*

$$(e_1^T \otimes I)\, Y_\ell = v_1 A_{k-\ell} - v_{\ell+1} A_k \quad \textit{for} \quad 1 \le \ell \le k-1\,. \tag{4.2.5}$$

Proof. Since $L(\lambda) \in \mathbb{L}_1(P)$, we know from Lemma 2.1.4 that

$$X \boxplus\!\!\rightarrow Y = v \otimes \begin{bmatrix} A_k & A_{k-1} & \dots & A_0 \end{bmatrix}. \tag{4.2.6}$$

From the definition of $\boxplus\!\!\rightarrow$ this implies that $X_1 = v \otimes A_k$, $Y_k = v \otimes A_0$, and

$$X_\ell + Y_{\ell-1} = v \otimes A_{k-\ell+1} \quad \text{for} \quad 2 \le \ell \le k\,, \tag{4.2.7}$$

Combining this with the convention $Y_0 := 0$, we then have (4.2.2) and (4.2.3), from which (4.2.4) immediately follows.

Since $L(\lambda) \in \mathbb{L}_2(P)$, we know from Lemma 2.1.11 that

$$X \boxplus\!\!\downarrow Y = v^T \otimes \begin{bmatrix} A_k \\ \vdots \\ A_0 \end{bmatrix} . \tag{4.2.8}$$

To extract from (4.2.8) a relation compatible with (4.2.3), we use the block-shift operator $N \otimes I$. Note that pre-multiplication of a block-column X_ℓ by $N \otimes I$ has the effect of shifting the blocks of X_ℓ up by one and losing the topmost block. Now if we slice off the topmost block-row of (4.2.8), then from the definition of $\boxplus\!\!\downarrow$ (expressed column-wise) we see that

$$(N \otimes I)X_\ell \,+\, Y_\ell \;=\; v_\ell \begin{bmatrix} A_{k-1} \\ \vdots \\ A_0 \end{bmatrix} \quad \text{for} \quad 1 \le \ell \le k\,. \tag{4.2.9}$$

Solving (4.2.9) for Y_ℓ and substituting in (4.2.3) for X_ℓ yields (4.2.1).

Finally, to establish (4.2.5) first rewrite (4.2.7) in the form

$$Y_\ell \;=\; (v \otimes A_{k-\ell}\,) \;-\; X_{\ell+1} \quad \text{for} \quad 1 \le \ell \le k-1\,.$$

Then skim off the topmost block by pre-multiplying by $e_1^T \otimes I$:

$$(e_1^T \otimes I)\,Y_\ell \;=\; (e_1^T v \otimes A_{k-\ell}\,) - (e_1^T \otimes I)X_{\ell+1} \;=\; v_1 A_{k-\ell} \;-\; (e_1^T \otimes I)X_{\ell+1}\,.$$

But (4.2.8) implies that the first row of X must be $v^T \otimes A_k$. Thus $(e_1^T \otimes I)X_{\ell+1} = v_{\ell+1}A_k$, and hence $(e_1^T \otimes I)\,Y_\ell \;=\; v_1 A_{k-\ell} \;-\; v_{\ell+1}A_k\,$. ∎

Using (4.2.1) we can now develop a concise formula describing the action of the block-row $\Lambda^T(x)\otimes I$ on the block-column Y_ℓ , where x is a scalar variable taking values in $\mathbb{C}$ and $\Lambda^T(x) := \begin{bmatrix} x^{k-1} & x^{k-2} & \dots & x & 1 \end{bmatrix}$. This formula will be used repeatedly and plays a central role in the proof of Theorem 4.2.6. (Note that $\Lambda^T(x)v$ is the same as the scalar v-polynomial $\mathsf{p}(x\,;v)$.)

Lemma 4.2.5. *Suppose that $L(\lambda) \in \mathbb{DL}(P)$ with ansatz vector v, and $\mathsf{p}(x\,;v)$ is the v-polynomial of v. Let Y_ℓ denote the ℓth block column of Y in $L(\lambda) = \lambda X + Y$, where $1 \le \ell \le k-1$. Then*

$$\Big(\Lambda^T(x) \otimes I\Big)Y_\ell \;=\; \mathsf{p}_{\ell-1}(x\,;v)P(x) - x\,\mathsf{p}(x\,;v)P_{\ell-1}(x), \tag{4.2.10}$$

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

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

$\boxed{\ell = 1}$: Using (4.2.1) we have

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

Simplifying this gives

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

which establishes (4.2.10) for $\ell = 1$. The induction hypothesis is now

$$\Big(\varLambda^T(x) \otimes I\Big)\, Y_{\ell-1} \;=\; \mathsf{p}_{\ell-2}(x\,;v)P(x) - x\,\mathsf{p}(x\,;v)P_{\ell-2}(x) \;. \tag{4.2.11}$$

$\boxed{\ell - 1 \,\Rightarrow\, \ell}$: Starting again with (4.2.1), we have

$$\begin{aligned}
\Big(\varLambda^T(x) \otimes I\Big)\, Y_\ell \;&=\; \Big(\varLambda^T(x) \otimes I\Big) \left((N \otimes I)(Y_{\ell-1} - v \otimes A_{k-\ell+1}) + v_\ell \begin{bmatrix} A_{k-1} \\ \vdots \\ A_0 \end{bmatrix} \right) \\
&=\; \Big(\varLambda^T(x) N \otimes I\Big)\big(Y_{\ell-1} - v \otimes A_{k-\ell+1}\big) + v_\ell \Big(\varLambda^T(x) \otimes I\Big) \begin{bmatrix} A_{k-1} \\ \vdots \\ A_0 \end{bmatrix} \\
&=\; \Big(\big(x\varLambda^T(x) - x^k e_1^T\big) \otimes I\Big)\big(Y_{\ell-1} - v \otimes A_{k-\ell+1}\big) + v_\ell\Big(P(x) - x^k A_k\Big) \\
&=\; x\Big(\varLambda^T(x) \otimes I\Big)Y_{\ell-1} - x^k\big(e_1^T \otimes I\big)Y_{\ell-1} - \big(x\varLambda^T(x)v\big)A_{k-\ell+1} \\
&\qquad + v_1 x^k A_{k-\ell+1} + v_\ell\, P(x) - v_\ell\, x^k A_k \;.
\end{aligned}$$

Note that $\big(e_1^T \otimes I\big)Y_{\ell-1}$ is the topmost block in $Y_{\ell-1}$, and is equal to $v_1 A_{k-\ell+1} - v_\ell A_k$ by (4.2.5). Finally, invoking the induction hypothesis (4.2.11) gives

$$\begin{aligned}
\Big(\varLambda^T(x) \otimes I\Big)\, Y_\ell \;&=\; x\,\mathsf{p}_{\ell-2}(x\,;v)P(x) - x^2\,\mathsf{p}(x\,;v)P_{\ell-2}(x) - v_1 x^k A_{k-\ell+1} + v_\ell\, x^k A_k \\
&\qquad - x\,\mathsf{p}(x\,;v)A_{k-\ell+1} + v_1 x^k A_{k-\ell+1} + v_\ell\, P(x) - v_\ell\, x^k A_k \\
&=\; \Big(x\,\mathsf{p}_{\ell-2}(x\,;v) + v_\ell\Big)P(x) - x\,\mathsf{p}(x\,;v)\Big(xP_{\ell-2}(x) + A_{k-\ell+1}\Big) \\
&=\; \mathsf{p}_{\ell-1}(x\,;v)P(x) - x\,\mathsf{p}(x\,;v)P_{\ell-1}(x) \;,
\end{aligned}$$

which completes the proof. $\square$

Theorem 4.2.6 (Determinant formula for pencils in $\mathbb{DL}(P)$).
Suppose that $L(\lambda)$ is in $\mathbb{DL}(P)$ with nonzero ansatz vector $v = [v_1, v_2, \dots, v_k]^T$. Assume that v has m leading zeroes with $0 \le m \le k-1$, so that $v_1 = v_2 = \cdots = v_m = 0$, $v_{m+1} \neq 0$ is the first nonzero coefficient of $\mathsf{p}(x\,; v)$, and $\mathsf{p}(x\,; v)$ has $k-m-1$ finite roots in $\mathbb{C}$, counted with multiplicities, denoted here by $r_1, r_2, \dots, r_{k-m-1}$. Then we have

$$\det L(\lambda) = \begin{cases} (-1)^{n\cdot\lfloor\frac{k}{2}\rfloor}(v_1)^{kn}\det\Big(P(r_1)P(r_2)\cdots P(r_{k-1})\Big)\det P(\lambda) & \text{if } m=0\,, \\ (-1)^s(v_{m+1})^{kn}(\det A_k)^m\det\Big(P(r_1)\cdots P(r_{k-m-1})\Big)\det P(\lambda) & \text{if } m>0\,, \end{cases} \tag{4.2.12}$$

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

Proof. The proof proceeds in three parts.

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

We begin the reduction process by right-multiplying $L(\lambda)$ by the block-Toeplitz matrix $T(\lambda)$. Recall that $T(\lambda)$ and $G(\lambda)$ denote the unimodular matrix polynomials defined in (1.1.6), and are related to each other via the factorization in (1.1.7). Using (4.2.4) for the description of $L(\lambda)$, an argument very similar to the one used in the proof of Theorem 2.2.1 yields the block-column-wise description

$$L(\lambda)G(\lambda) = \left[\begin{array}{c|c|c|c|c} Y_1 + \lambda v\otimes A_k & Y_2 - \lambda Y_1 + \lambda\, v\otimes A_{k-1} & \cdots & Y_{k-1} - \lambda Y_{k-2} + \lambda\, v\otimes A_2 & v\otimes P(\lambda) \end{array}\right],$$

and hence

$$L(\lambda)T(\lambda) = \left[\begin{array}{c|c|c|c|c} Y_1 + \lambda v\otimes P_0(\lambda) & Y_2 + \lambda v\otimes P_1(\lambda) & \cdots & Y_{k-1} + \lambda v\otimes P_{k-2}(\lambda) & v\otimes P(\lambda) \end{array}\right]. \tag{4.2.13}$$

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

$$M := \begin{bmatrix} e_1^T \\ \Lambda^T(r_1) \\ \Lambda^T(r_2) \\ \vdots \\ \Lambda^T(r_{k-1}) \end{bmatrix} \otimes I = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ r_1^{k-1} & r_1^{k-2} & \cdots & r_1 & 1 \\ r_2^{k-1} & r_2^{k-2} & \cdots & r_2 & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ r_{k-1}^{k-1} & r_{k-1}^{k-2} & \cdots & r_{k-1} & 1 \end{bmatrix} \otimes I\,. \tag{4.2.14}$$

Using Lemma 4.2.5 and the fact that $\Lambda^T(r_j)v = \mathsf{p}(r_j\,; v)$, we obtain that

$$\begin{aligned}&\Big(\Lambda^T(r_j)\otimes I\Big)\Big(Y_\ell + \lambda v\otimes P_{\ell-1}(\lambda)\Big) \\ &\qquad = \mathsf{p}_{\ell-1}(r_j\,; v)P(r_j) - r_j\,\mathsf{p}(r_j\,; v)P_{\ell-1}(r_j) + \lambda\,\mathsf{p}(r_j\,; v)P_{\ell-1}(\lambda)\,.\end{aligned}$$

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

$$\left[\begin{array}{cccc|c} * & * & \cdots & * & v_1P(\lambda) \\ \hline \mathsf{p}_0(r_1\,;v)P(r_1) & \mathsf{p}_1(r_1\,;v)P(r_1) & \cdots & \mathsf{p}_{k-2}(r_1\,;v)P(r_1) & 0 \\ \mathsf{p}_0(r_2\,;v)P(r_2) & \mathsf{p}_1(r_2\,;v)P(r_2) & \cdots & \mathsf{p}_{k-2}(r_2\,;v)P(r_2) & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathsf{p}_0(r_{k-1}\,;v)P(r_{k-1}) & \mathsf{p}_1(r_{k-1}\,;v)P(r_{k-1}) & \cdots & \mathsf{p}_{k-2}(r_{k-1}\,;v)P(r_{k-1}) & 0 \end{array}\right].$$

This matrix now factors into

$$\underbrace{\left[\begin{array}{c|ccc} I & & & \\ \hline & P(r_1) & & \\ & & \ddots & \\ & & & P(r_{k-1}) \end{array}\right]}_{=:W} \left[\begin{array}{ccc|c} * & \cdots & * & v_1P(\lambda) \\ \hline \mathsf{p}_0(r_1\,;v)I & \cdots & \mathsf{p}_{k-2}(r_1\,;v)I & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \mathsf{p}_0(r_{k-1}\,;v)I & \cdots & \mathsf{p}_{k-2}(r_{k-1}\,;v)I & 0 \end{array}\right],$$

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

$$ML(\lambda)T(\lambda)(R\otimes I) \;=\; W \left[\begin{array}{c|c} v_1P(\lambda) & * \\ \hline 0 & \\ \vdots & V\otimes I \\ 0 & \end{array}\right], \tag{4.2.15}$$

where

$$\begin{aligned} V \;&=\; \begin{bmatrix} \mathsf{p}_{k-2}(r_1\,;v) & \cdots & \mathsf{p}_1(r_1\,;v) & \mathsf{p}_0(r_1\,;v) \\ \vdots & \vdots & \vdots & \vdots \\ \mathsf{p}_{k-2}(r_{k-1}\,;v) & \cdots & \mathsf{p}_1(r_{k-1}\,;v) & \mathsf{p}_0(r_{k-1}\,;v) \end{bmatrix} \\ &=\; \begin{bmatrix} (v_1r_1^{k-2} + \cdots + v_{k-2}r_1 + v_{k-1}) & \cdots & (v_1r_1 + v_2) & v_1 \\ \vdots & \vdots & \vdots & \vdots \\ (v_1r_{k-1}^{k-2} + \cdots + v_{k-2}r_{k-1} + v_{k-1}) & \cdots & (v_1r_{k-1} + v_2) & v_1 \end{bmatrix}. \end{aligned}$$

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

$$v_1 \cdot \begin{bmatrix} r_1^{k-2} & r_1^{k-3} & \cdots & r_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{k-1}^{k-2} & r_{k-1}^{k-3} & \cdots & r_{k-1} & 1 \end{bmatrix}, \tag{4.2.16}$$

so $\det(V\otimes I) = v_1^{(k-1)n}\det M$. Taking determinants on both sides of (4.2.15) now gives

$$\begin{aligned} &\det M \cdot \det L(\lambda) \cdot \det T(\lambda) \cdot \det(R\otimes I) \\ &\qquad\qquad = \det\Big(P(r_1)P(r_2)\cdots P(r_{k-1})\Big) \cdot \det\big(v_1P(\lambda)\big) \cdot \det(V\otimes I)\,. \end{aligned}$$

Since

$$\det(R\otimes I) = \det(R_k\otimes I_n) = (\det R_k)^n(\det I_n)^k = (-1)^{n\cdot\lfloor\frac{k}{2}\rfloor} \tag{4.2.17}$$

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

$$\det L(\lambda) \;=\; (-1)^{n\cdot\lfloor\frac{k}{2}\rfloor}(v_1)^{kn}\det\Big(P(r_1)P(r_2)\cdots P(r_{k-1})\Big)\det P(\lambda)\,. \tag{4.2.18}$$

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

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

We start out the same way as before, postmultiplying $L(\lambda)$ by $T(\lambda)$ to get (4.2.13). But then, instead of M in (4.2.14), we use all available finite roots of $\mathsf{p}(x\,;v)$ to define the following modified version of M:

$$\widehat{M} \;:=\; \begin{bmatrix} e_1^T \\ \vdots \\ e_{m+1}^T \\ \Lambda^T(r_1) \\ \vdots \\ \Lambda^T(r_{k-m-1}) \end{bmatrix} \otimes I_n \;=\; \left[\begin{array}{ccccc} & I_{m+1} & & 0 & \\ \hline r_1^{k-1} & r_1^{k-2} & \cdots & r_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{k-m-1}^{k-1} & r_{k-m-1}^{k-2} & \cdots & r_{k-m-1} & 1 \end{array}\right] \otimes I_n\,. \tag{4.2.19}$$

Now simplify the product $\widehat{M}L(\lambda)T(\lambda)$ using Lemma 4.2.5 and $\Lambda^T(r_\ell)v = \mathsf{p}(r_\ell\,;v) = 0$ as before, as well as the fact that $v_1 = v_2 = \cdots = v_m = 0$, which implies that $\mathsf{p}_0(x\,;v)$, $\mathsf{p}_1(x\,;v)$, $\ldots$, $\mathsf{p}_{m-1}(x\,;v)$ are all zero polynomials. Then we obtain

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

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

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

where the $mn \times mn$ block B can also be seen to have some further structure. First note that because of the structure of $\widehat{M}$, the block B in $\widehat{M}L(\lambda)T(\lambda)$ is exactly the same as the corresponding block in $L(\lambda)T(\lambda)$ in (4.2.13), which is just the first mn rows of

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

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

$$\left[\begin{array}{c|c|c|c} Y_1 & Y_2 & \cdots & Y_m \end{array}\right].$$

Using the recurrence (4.2.1) from Lemma 4.2.4 with $1 \leq \ell \leq m$, we can now show that B is actually block anti-triangular. When $\ell = 1$ we have $Y_1 = -Nv \otimes A_k$. Since the first m entries of Nv are $[v_2, v_3, \ldots, v_{m+1}]^T = [0, 0, \ldots, v_{m+1}]^T$, we see that the first block-column of B is $\left[\begin{array}{cccc} 0, & \ldots, & 0, & -v_{m+1}A_k^T \end{array}\right]^T$. With $\ell = 2$ we have $Y_2 = (N \otimes I)Y_1 - Nv \otimes A_{k-1}$, whose first mn rows are

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

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

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

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

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

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

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

Performing some block-column permutations gives us

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

$$\left[\begin{array}{cc|c|ccc}
-v_{m+1}A_k & 0 & 0 & & & \\
\ddots & & \vdots & & * & \\
* & -v_{m+1}A_k & 0 & & & \\
\hline
\multicolumn{2}{c|}{*} & v_{m+1}P(\lambda) & * & \cdots & * \\
\hline
 & & 0 & \mathsf{p}_{k-2}(r_1\,;v)P(r_1) & \cdots & \mathsf{p}_m(r_1\,;v)P(r_1) \\
\multicolumn{2}{c|}{0} & \vdots & \vdots & \vdots & \vdots \\
 & & 0 & \mathsf{p}_{k-2}(r_{k-m-1}\,;v)P(r_{k-m-1}) & \cdots & \mathsf{p}_m(r_{k-m-1}\,;v)P(r_{k-m-1})
\end{array}\right], \tag{4.2.20}$$

which after factoring becomes

$$\left[\begin{array}{c|c|c}
(-v_{m+1}I_m)\otimes I_n & 0 & 0 \\
\hline
0 & v_{m+1}I_n & 0 \\
\hline
0 & 0 & \widehat{W}
\end{array}\right]
\left[\begin{array}{cc|c|c}
A_k & 0 & & \\
\ddots & & 0 & * \\
* & A_k & & \\
\hline
\multicolumn{2}{c|}{0} & P(\lambda) & * \\
\hline
\multicolumn{2}{c|}{0} & 0 & \widehat{V}\otimes I_n
\end{array}\right], \tag{4.2.21}$$

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

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

Since $v_{m+1} \neq 0$, this $(k-m-1)\times(k-m-1)$ matrix $\widehat{V}$ can be reduced by $(\det = +1)$ column operations in a manner analogous to the reduction of V in (4.2.16), so we see that

$$\det(\widehat{V} \otimes I_n) \;=\; (v_{m+1})^{(k-m-1)n} \det \widehat{M}\,. \tag{4.2.22}$$

Now taking determinants on both sides of (4.2.20) using the factorization (4.2.21) gives

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

Cancelling $\det \widehat{M}$ on both sides using (4.2.22), and using $\det T(\lambda) = +1$ together with the fact that $\det(R \otimes I)$ is its own inverse, we get

$$\begin{aligned}
\det L(\lambda) \;=\; &\det\Big(P(r_1)P(r_2)\cdots P(r_{k-m-1})\Big) \cdot (-1)^{mn} \cdot (v_{m+1})^{kn} \cdot (\det A_k)^m \\
&\cdot \det P(\lambda) \cdot \det(R_m \otimes I_n) \cdot \det(R_{k-m} \otimes I_n)\,.
\end{aligned}$$

Finally, substituting $\det(R_m \otimes I_n) = (-1)^{n \cdot \lfloor \frac{m}{2} \rfloor}$ and $\det(R_{k-m} \otimes I_n) = (-1)^{n \cdot \lfloor \frac{k-m}{2} \rfloor}$ from (4.2.17) yields the desired formula (4.2.12). Note that this is consistent with formula (4.2.18) derived for the $m = 0$ case, as long as we interpret the term $(\det A_k)^m$ to be equal to $+1$ whenever $m = 0$, regardless of whether $\det A_k$ is zero or nonzero.

Part 3 : Now that we know that (4.2.12) holds for any $v \in \mathbb{F}^k$ such that the corresponding $\mathsf{p}(x\,; v)$ has *distinct* finite roots, we can leverage this result to the general case by a continuity argument. For every *fixed* m and fixed polynomial $P(\lambda)$, the formula on the right-hand side of (4.2.12) is clearly a continuous function of the leading coefficient v_{m+1} and the roots $r_1, r_2, \ldots, r_{k-m-1}$ of $\mathsf{p}(x\,; v)$, and is defined for all lists in the set $\mathcal{D} = \{(v_{m+1}, r_1, r_2, \ldots, r_{k-m-1}) : v_{m+1} \neq 0\}$, regardless of whether the numbers $r_1, r_2, \ldots, r_{k-m-1}$ are distinct or not.

The left-hand side of (4.2.12) can also be viewed as a function defined and continuous for all lists in $\mathcal{D}$. To see this, first observe that the map

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

taking the leading coefficient and roots of the polynomial $\mathsf{p}(x\,; v)$ to the coefficients of the same polynomial $\mathsf{p}(x\,; v)$ is defined and continuous on $\mathcal{D}$, as well as being surjective. Next note that because of Theorem 3.4.2 and the isomorphism in (3.3.1), the unique pencil $L(\lambda) \in \mathbb{DL}(P)$ corresponding to $v = (0, 0, \ldots, 0, v_{m+1}, \ldots, v_k)^T$ can be expressed as a linear combination

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

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

In summary, the two sides of (4.2.12) are continuous functions defined on the same domain $\mathcal{D}$, and have been shown to be equal on a *dense* subset

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

of $\mathcal{D}$. Therefore by continuity the two sides of (4.2.12) must be equal on all of $\mathcal{D}$. Since this argument holds for each fixed m with $0 \leq m \leq k-1$, the desired result is established for all nonzero $v \in \mathbb{F}^k$. $\square$

4.3 The Eigenvalue Exclusion Theorem

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

Theorem 4.3.1 (Eigenvalue Exclusion Theorem).
Suppose that $P(\lambda)$ is a regular matrix polynomial and $L(\lambda)$ is in $\mathbb{DL}(P)$ with nonzero ansatz vector v. Then $L(\lambda)$ is a (strong) linearization for $P(\lambda)$ if and only if no root of the v*-polynomial $\mathsf{p}(x\,; v)$ is an eigenvalue of $P(\lambda)$. (Note that this statement includes ∞ as one of the possible roots of $\mathsf{p}(x\,; v)$ or possible eigenvalues of $P(\lambda)$.)*

Proof. By Theorem 2.2.3, $L(\lambda)$ is a (strong) linearization for $P(\lambda)$ if and only if $L(\lambda)$ is regular. But from the determinant formula (4.2.12) it follows that $L(\lambda)$ is regular if and only if no root of $\mathsf{p}(x\,;v)$ is an eigenvalue of $P(\lambda)$. ∎

Using Theorem 4.3.1 we can now show that almost every pencil in $\mathbb{DL}(P)$ is a linearization for P. Although the same property was proved in Theorem 2.2.7 for pencils in $\mathbb{L}_1(P)$, the result for $\mathbb{DL}(P)$ is not a consequence of Theorem 2.2.7, since $\mathbb{DL}(P)$ is itself a closed, nowhere dense subset of measure zero in $\mathbb{L}_1(P)$. Neither can the proof of Theorem 2.2.7 be directly generalized in any simple way; hence the need for a different argument in the following result.

Theorem 4.3.2 (Linearizations are Generic[1] in $\mathbb{DL}(P)$).
For any regular matrix polynomial $P(\lambda)$, pencils in $\mathbb{DL}(P)$ are linearizations of $P(\lambda)$ for almost all $v \in \mathbb{F}^k$. (Here "almost all" means for all but a closed, nowhere dense set of measure zero in $\mathbb{F}^k$.)

Proof. Recall that the resultant $\operatorname{res}(f,g)$ of two polynomials $f(x)$ and $g(x)$ is a polynomial in the *coefficients* of f and g with the property that $\operatorname{res}(f,g) = 0$ if and only if $f(x)$ and $g(x)$ have a common(finite) root [78, p.248], [84]. Now consider $r = \operatorname{res}\Big(\mathsf{p}(x\,;v), \det P(x)\Big)$, which, because $P(\lambda)$ is fixed, can be viewed as a polynomial $r(v_1, v_2, \ldots, v_k)$ in the components of $v \in \mathbb{F}^k$. The zero set $\mathcal{Z}(r) = \big\{v \in \mathbb{F}^k : r(v_1, v_2, \ldots, v_k) = 0\big\}$, then, is exactly the set of $v \in \mathbb{F}^k$ for which some finite root of $\mathsf{p}(x\,;v)$ is an eigenvalue of $P(\lambda)$, together with the point $v = 0$. To see that $\mathcal{Z}(r)$ is always a *proper* subset of $\mathbb{F}^k$, it suffices to find some $v \in \mathbb{R}^k$ such that $r(v) \neq 0$, i.e. such that $\mathsf{p}(x\,;v)$ has no (finite) root that is an eigenvalue of P. Here is a way to find many such v: pick $v = e_1 - \alpha e_k$ with sufficiently large $\alpha > 0$ so that all roots of the corresponding v-polynomial $\mathsf{p}(x\,;v) = x^{k-1} - \alpha$ are bigger in modulus than all the finite eigenvalues of $P(\lambda)$.

Now recall our convention that the v-polynomial $\mathsf{p}(x\,;v)$ has ∞ as a root exactly for $v \in \mathbb{F}^k$ lying in the hyperplane $v_1 = 0$. Then by Theorem 4.3.1 the set of vectors $v \in \mathbb{F}^k$ for which the corresponding pencil $L(\lambda) \in \mathbb{DL}(P)$ is *not* a linearization[2] of $P(\lambda)$ is either the proper algebraic set $\mathcal{Z}(r)$ or the union of two proper algebraic sets, $\mathcal{Z}(r)$ and the hyperplane $v_1 = 0$. But the union of any finite number of proper algebraic sets is always a closed, nowhere dense set of measure zero in $\mathbb{F}^k$. ∎

How far can the eigenvalue exclusion theorem be extended from $\mathbb{DL}(P)$-pencils to other pencils in $\mathbb{L}_1(P)$? Let us say that a pencil $L \in \mathbb{L}_1(P)$ with right ansatz vector v has the *eigenvalue exclusion property* if the statement "no root of the v-polynomial $\mathsf{p}(x\,;v)$ is an eigenvalue of $P(\lambda)$" is equivalent to the linearization condition for L. That there are pencils in $\mathbb{L}_1(P)$ with the eigenvalue exclusion property that are not in $\mathbb{DL}(P)$ is shown by the pencil $L_1(\lambda)$ in Example 2.2.5. The following variation of Example 2.2.6, though, is easily shown not to have the eigenvalue exclusion property.

Example 4.3.3. *For the general cubic polynomial $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$*

[1]Because of Theorem 3.4.2 this property also, of course, holds for $\mathbb{B}(P)$.

[2]Note that $v = 0$ corresponds to the zero pencil $L(\lambda)$ in $\mathbb{DL}(P)$, which is never a linearization.

consider the pencil

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

that is in $\mathbb{L}_1(P)$ *but not in* $\mathbb{DL}(P)$. *Since* $X \boxplus\!\!\rightarrow Y = \begin{bmatrix} 1 & -2 & 0 \end{bmatrix}^T \otimes \begin{bmatrix} A & B & C & D \end{bmatrix}$, *the right ansatz vector is* $v = \begin{bmatrix} 1 & -2 & 0 \end{bmatrix}^T$ *with* v*-polynomial* $\mathsf{p}(x\,; v) = x^2 - 2x$ *and roots* 0 *and* 2. *On the other hand, applying the procedure described in Section* 2.2.2 *gives*

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

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

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

Chapter 5

Symmetric and Hermitian Linearizations

We now return to one of the problems that originally motivated the investigation in this thesis, that of systematically finding large sets of symmetric linearizations for symmetric polynomials, $P(\lambda) = P(\lambda)^T$. Our strategy is first to characterize the subspace

$$\mathbb{S}(P) := \left\{ \lambda X + Y \in \mathbb{L}_1(P) : X^T = X,\, Y^T = Y \right\} \tag{5.0.1}$$

of all symmetric pencils in $\mathbb{L}_1(P)$ when P is symmetric, and then later in Section 5.3 show that almost all of these symmetric pencils are indeed linearizations for P. An analogous development for Hermitian P is carried out in Section 5.2.

5.1 Symmetric Pencils in $\mathbb{L}_1(P)$ for Symmetric P

We begin with a result for symmetric polynomials reminiscent of Theorem 3.1.6, but using transpose rather than block transpose.

Lemma 5.1.1. *Suppose $P(\lambda)$ is a symmetric matrix polynomial and $L(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector v. Then $L^T(\lambda) \in \mathbb{L}_2(P)$ with left ansatz vector $w = v$. Similarly, $L(\lambda) \in \mathbb{L}_2(P)$ with left ansatz vector v implies that $L^T(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector v.*

Proof. Suppose $L(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector v. Then

$$\Big(L(\lambda)(\Lambda \otimes I)\Big)^T = \Big(v \otimes P(\lambda)\Big)^T \quad \Longrightarrow \quad (\Lambda^T \otimes I)L^T(\lambda) \;=\; v^T \otimes P(\lambda).$$

Thus $L^T(\lambda) \in \mathbb{L}_2(P)$ with left ansatz vector v. The proof of the second statement is analogous. ∎

The space $\mathbb{S}(P)$ is characterized in the next result by relating it to the previously developed space $\mathbb{DL}(P)$.

Theorem 5.1.2 (Characterization of $\mathbb{S}(P)$).
For any symmetric polynomial $P(\lambda)$, $\mathbb{S}(P) = \mathbb{DL}(P)$.

Proof. Suppose $L(\lambda) \in \mathbb{S}(P) \subseteq \mathbb{L}_1(P)$ with right ansatz vector v. Then by Lemma 5.1.1 we know that $L^T(\lambda) = L(\lambda)$ is in $\mathbb{L}_2(P)$ with left ansatz vector v, and so $L(\lambda) \in \mathbb{DL}(P)$. Thus $\mathbb{S}(P) \subseteq \mathbb{DL}(P)$.

By Lemma 5.1.1, $L(\lambda) \in \mathbb{DL}(P)$ with right/left ansatz vector v implies that $L^T(\lambda) \in \mathbb{DL}(P)$ with left/right ansatz vector v. But by Theorem 3.4.2 pencils in $\mathbb{DL}(P)$ are uniquely determined by their ansatz vector, so $L(\lambda) \equiv L^T(\lambda)$, and hence $\mathbb{DL}(P) \subseteq \mathbb{S}(P)$. Therefore $\mathbb{DL}(P) = \mathbb{S}(P)$. ∎

Once again one may refer to Tables 3.3.1 and 3.3.2 for examples of what are in effect *triply*-structured pencils whenever P is symmetric. Recall, however, that there are symmetric linearizations for P that are not in $\mathbb{S}(P)$: L in (3.5.1) is not in $\mathbb{S}(P)$, but is a symmetric linearization for any symmetric cubic P.

5.2 Hermitian Pencils in $\mathbb{L}_1(P)$ for Hermitian P

For a Hermitian matrix polynomial $P(\lambda)$ of degree k, that is, $P(\lambda)^* = P(\overline{\lambda})$, let

$$\mathbb{H}(P) := \left\{ \lambda X + Y \in \mathbb{L}_1(P) : X^* = X,\ Y^* = Y \right\} \tag{5.2.1}$$

denote the set of all Hermitian pencils in $\mathbb{L}_1(P)$. A priori the right ansatz vector v of a pencil in $\mathbb{H}(P)$ might be any vector in $\mathbb{C}^k$, since P is a complex polynomial. However, the next result shows that any such v must in fact be real.

Lemma 5.2.1. *Suppose $P(\lambda)$ is a Hermitian polynomial and $L(\lambda) \in \mathbb{H}(P)$ with right ansatz vector v. Then $v \in \mathbb{R}^k$ and $L(\lambda) \in \mathbb{DL}(P)$, so $\mathbb{H}(P) \subsetneq \mathbb{DL}(P)$.*

Proof. Since $L(\lambda) \in \mathbb{L}_1(P)$, we have $L(\lambda)(\Lambda \otimes I) = v \otimes P(\lambda)$. Then, since P and L are Hermitian,

$$\Big(L(\lambda)(\Lambda \otimes I)\Big)^* = \Big(v \otimes P(\lambda)\Big)^* \quad \Longrightarrow \quad (\overline{\Lambda}^T \otimes I)L(\overline{\lambda}) = \overline{v}^T \otimes P(\overline{\lambda}).$$

This last equation holds for all λ, so we may replace $\overline{\lambda}$ by λ to get $(\Lambda^T \otimes I) \cdot L(\lambda) = \overline{v}^T \otimes P(\lambda)$, so that $L(\lambda) \in \mathbb{L}_2(P)$ with left ansatz vector $w = \overline{v}$. Thus $L(\lambda) \in \mathbb{DL}(P)$. But by Theorem 3.4.2 the right and left ansatz vectors of any $\mathbb{DL}(P)$-pencil must be equal. So $v = \overline{v}$, which means $v \in \mathbb{R}^k$. Since $\mathbb{DL}(P)$ includes pencils corresponding to nonreal v, $\mathbb{H}(P) \subsetneq \mathbb{DL}(P)$. ∎

Recall the map $\mathbb{DL}(P) \xrightarrow{\mathcal{M}} \mathcal{V}_P$ from (3.3.1), which we know from Theorem 3.3.1 and Theorem 3.4.2 to be an isomorphism. Lemma 5.2.1 implies that $\mathcal{M}$ can be restricted to the subspace $\mathbb{H}(P)$, giving a 1-1 map into the "real" part of $\mathcal{V}_P$, i.e. into the subspace $\mathcal{R}_P := \left\{ v \otimes P(\lambda) : v \in \mathbb{R}^k \right\} \subsetneq \mathcal{V}_P$. The characterization of $\mathbb{H}(P)$ is then completed in the next result by showing that $\mathbb{H}(P) \xrightarrow{\mathcal{M}} \mathcal{R}_P$ is actually an isomorphism[1].

Theorem 5.2.2 (Characterization of $\mathbb{H}(P)$).
For any Hermitian polynomial $P(\lambda)$, $\mathbb{H}(P)$ is the subset of all pencils in $\mathbb{DL}(P)$ with a real ansatz vector. In other words, for each vector $v \in \mathbb{R}^k$ there is a unique Hermitian pencil $H(\lambda) \in \mathbb{H}(P)$.

[1] Note that $\mathbb{H}(P)$ is only a *real* subspace of $\mathbb{DL}(P)$, so this is an isomorphism of real vector spaces.

Proof. We need to show that the map $\mathbb{H}(P) \xrightarrow{\mathcal{M}} \mathcal{R}_P$ is an isomorphism, and from the remarks preceding the theorem all that remains is to show that the map $\mathcal{M}$ is onto. By arguments analogous to the ones used in Lemma 5.1.1 and Theorem 5.1.2, it is straightforward to show that for Hermitian P, $L(\lambda) \in \mathbb{DL}(P)$ with right/left ansatz vector v implies that $L^*(\lambda) \in \mathbb{DL}(P)$ with left/right ansatz vector $\overline{v}$. Now if for an arbitrary $v \in \mathbb{R}^k$ we let $H(\lambda)$ be the unique pencil in $\mathbb{DL}(P)$ with right/left ansatz vector v, then $H^*(\lambda)$ is also in $\mathbb{DL}(P)$ with exactly the same ansatz vector v. The uniqueness of $\mathbb{DL}(P)$-pencils then implies that we must have $H(\lambda) \equiv H^*(\lambda)$, i.e., $H(\lambda) \in \mathbb{H}(P)$, thus showing that the map $\mathcal{M}$ is onto. ∎

5.3 Genericity of Linearizations in $\mathbb{S}(P)$ and $\mathbb{H}(P)$

The remaining basic issue regarding $\mathbb{S}(P)$ and $\mathbb{H}(P)$ is the question of which pencils in these spaces are actually linearizations for P, when P is symmetric or Hermitian, respectively. The theory developed thus far provides us with several ways to address this question. First of all, the Strong Linearization Theorem (Thm. 2.2.3) tells us that for regular P, a pencil L in $\mathbb{S}(P)$ or $\mathbb{H}(P)$ is a linearization precisely when L itself is regular. Secondly, for any given pencil in $\mathbb{S}(P)$ or $\mathbb{H}(P)$ the procedure described in Section 2.2.2 allows one to derive an individualized symbolic "linearization condition" which determines whether the given pencil is a linearization or not. Finally, since any pencil in $\mathbb{S}(P)$ or $\mathbb{H}(P)$ is also in $\mathbb{DL}(P)$, the eigenvalue exclusion theorem from Chapter 4 applies, and gives yet a third criterion for deciding if such a pencil is a linearization. We recall that result here for the convenience of the reader.

Theorem 4.3.1 (Eigenvalue Exclusion Theorem).
Suppose that $P(\lambda)$ is a regular matrix polynomial and $L(\lambda) \in \mathbb{DL}(P)$ with nonzero ansatz vector v. Then $L(\lambda)$ is a linearization for $P(\lambda)$ if and only if no root of the v*-polynomial* $\mathsf{p}(x\,;v)$ *is an eigenvalue of $P(\lambda)$.*

As a consequence of Theorem 4.3.1 we also showed in Chapter 4 that almost every pencil in $\mathbb{DL}(P)$ is a linearization, where "almost every" means all except for a closed, nowhere dense set of measure zero. Because $\mathbb{S}(P) = \mathbb{DL}(P)$ when P is symmetric (Thm. 5.1.2), the same result holds for $\mathbb{S}(P)$. However, when P is Hermitian the space $\mathbb{H}(P)$ is a closed, nowhere dense subset of measure zero[2] in $\mathbb{DL}(P)$, so we cannot immediately deduce an "almost every" result for $\mathbb{H}(P)$. Some further analysis is therefore required. It turns out that only small modifications to the argument for Theorem 4.3.2 are needed to prove the following result.

Theorem 5.3.1 (Linearizations are Generic in $\mathbb{H}(P)$).
Let $P(\lambda)$ be a regular Hermitian matrix polynomial. For almost every $v \in \mathbb{R}^k$ the corresponding pencil in $\mathbb{H}(P)$ is a linearization.

Proof. Consider again the resultant $r = \operatorname{res}\Big(\mathsf{p}(x\,;v),\, \det P(x)\Big)$, which, because P is fixed, can be viewed as a polynomial $r(v_1, v_2, \ldots, v_k)$ in the components of $v \in \mathbb{R}^k$. Since P is Hermitian, i.e. $P(\overline{x})^* = P(x)$, we have for $g(x) := \det P(x)$ that

$$g(x) = \det P(x) = \det\big(P(\overline{x})^*\big) = \overline{\det P(\overline{x})} = \overline{g(\overline{x})} = \overline{g}(x)\,.$$

[2] More precisely, the ansatz vector set $\mathbb{R}^k$ of $\mathbb{H}(P)$ is a closed, nowhere dense subset of measure zero in the ansatz vector set $\mathbb{C}^k$ of $\mathbb{DL}(P)$.

Thus all the coefficients of the polynomial $\det P(x)$ are real, and hence the resultant r is a *real* polynomial in the real variables $v_1, v_2, \ldots, v_k$. The real zero set $\mathcal{Z}(r) = \left\{ v \in \mathbb{R}^k : r(v_1, v_2, \ldots, v_k) = 0 \right\} \subseteq \mathbb{R}^k$ is exactly the set of all $v \in \mathbb{R}^k$ for which some finite root of $\mathsf{p}(x\,; v)$ is an eigenvalue of $P(\lambda)$, together with $v = 0$. By the argument given in Theorem 4.3.2 we know that $\mathcal{Z}(r)$ is a *proper* algebraic subset of $\mathbb{R}^k$.

Now recall our convention that the v-polynomial $\mathsf{p}(x\,; v)$ has ∞ as a root exactly for $v \in \mathbb{R}^k$ lying in the hyperplane $v_1 = 0$. Then by Theorem 4.3.1 the set of vectors $v \in \mathbb{R}^k$ for which the corresponding pencil $L(\lambda) \in \mathbb{H}(P) \subset \mathbb{DL}(P)$ is *not* a linearization of $P(\lambda)$ is either the proper (real) algebraic set $\mathcal{Z}(r)$, or the union of two proper (real) algebraic sets, $\mathcal{Z}(r)$ and the hyperplane $v_1 = 0$. But the union of any finite number of proper (real) algebraic sets is always a closed, nowhere dense set of measure zero in $\mathbb{R}^k$. ∎

Chapter 6

Palindromic and Alternating Polynomials

The previous chapter considered two types of structured matrix polynomial, symmetric and Hermitian, that are well known in applications to structural mechanics. There it was shown that $\mathbb{L}_1(P)$, or more specifically $\mathbb{DL}(P)$, is a rich source of structure-preserving linearizations for such polynomials.

Now we consider several other kinds of polynomial structure, which we broadly refer to as palindromic and alternating. These are rather less familiar than symmetric and Hermitian structure, but nevertheless important in a wide variety of applications. After giving precise definitions and some examples, we describe the spectral symmetries associated with these polynomial structures, and show how these structures are related to each other via a matrix polynomial analog of the well-known Cayley transformation of matrices. In the course of this development we will see a number of ways in which palindromic and alternating polynomials can reasonably be viewed as generalizations (or at least analogs) of symplectic and Hamiltonian matrices. Finally, we conclude the chapter with brief descriptions of several applications where palindromic or alternating polynomials occur.

6.1 Basic Structures: Definitions and Notation

One of the questions that originally motivated the investigation contained in this thesis is an eigenvalue problem arising in the study of rail traffic noise caused by high speed trains. This problem has the form

$$\big(\lambda^2 A + \lambda B + A^T\big)x = 0\,, \tag{6.1.1}$$

where A, B are complex square matrices with B complex symmetric and A singular. Observe that the matrix polynomial in (6.1.1) has the property that reversing the order of the coefficient matrices, followed by taking their transpose, leads back to the original matrix polynomial. By analogy with linguistic palindromes (see section 6.5) we have chosen to use the name *T-palindromic* for such a polynomial. Further details about this application may be found in Example 6.4.1 and the references [37], [38], and [42].

A different type of structured eigenvalue problem, arising in gyroscopic systems and in the study of elastic materials, has the form

$$(\lambda^2 M + \lambda G + K)x = 0\,, \tag{6.1.2}$$

where M and K are real symmetric while G is real *skew*-symmetric. The matrix polynomial in (6.1.2) is somewhat reminiscent of an even function: replacing λ by $-\lambda$ followed by taking the transpose returns us back to the original matrix polynomial. Hence we denote such matrix polynomials by the term T*-even*.

Variations of T-palindromic and T-even structure are also possible; for example, transpose T can be replaced by conjugate transpose $*$, or even left out altogether. It turns out that these two apparently very different kinds of structure are actually quite closely related, and it is for this reason that we study and develop their properties together, along with their variants defined below. Further details of applications involving these structures can be found in Section 6.4 and the references cited therein, as well as in [83].

We begin by defining two operations on matrix polynomials, $\star$-adjoint and reversal. For conciseness, the symbol $\star$ is used as an abbreviation for transpose T in the real case and either T or conjugate transpose $*$ in the complex case.

Definition 6.1.1 (Adjoint and Reversal of Matrix Polynomials).
Let $Q(\lambda) = \sum_{i=0}^{k} \lambda^i B_i$, where $B_0, \ldots, B_k \in \mathbb{F}^{m\times n}$, be a matrix polynomial of degree k, that is, $B_k \neq 0$. Then

$$Q^\star(\lambda) := \sum_{i=0}^{k} \lambda^i B_i^\star \quad \text{and} \quad \mathrm{rev}Q(\lambda) := \lambda^k Q(1/\lambda) = \sum_{i=0}^{k} \lambda^i B_{k-i} \tag{6.1.3}$$

defines the $\star$-*adjoint* $Q^\star(\lambda)$ and the *reversal* $\mathrm{rev}Q(\lambda)$ of $Q(\lambda)$, respectively.

If $\deg(Q(\lambda))$ denotes the *degree* of the matrix polynomial $Q(\lambda)$, then, in general, $\deg(\mathrm{rev}Q(\lambda)) \le \deg(Q(\lambda))$ and $\mathrm{rev}\big(Q_1(\lambda)\cdot Q_2(\lambda)\big) = \mathrm{rev}Q_1(\lambda)\cdot \mathrm{rev}Q_2(\lambda)$, whenever the product $Q_1(\lambda)\cdot Q_2(\lambda)$ is defined. Using the operations in (6.1.3), the various structured matrix polynomials to be considered are now defined in Table 6.1.1.

Table 6.1.1: Definitions of basic structures

palindromic	$\mathrm{rev}\,P(\lambda) = P(\lambda)$	anti-palindromic	$\mathrm{rev}\,P(\lambda) = -P(\lambda)$
$\star$-palindromic	$\mathrm{rev}\,P^\star(\lambda) = P(\lambda)$	$\star$-anti-palindromic	$\mathrm{rev}\,P^\star(\lambda) = -P(\lambda)$
even	$P(-\lambda) = P(\lambda)$	odd	$P(-\lambda) = -P(\lambda)$
$\star$-even	$P^\star(-\lambda) = P(\lambda)$	$\star$-odd	$P^\star(-\lambda) = -P(\lambda)$

For a scalar polynomial $p(x)$, being T-palindromic is the same as being palindromic (i.e., $\mathrm{rev}p(x) = p(x)$), while $*$-palindromic is equivalent to being conjugate-palindromic (i.e., $\mathrm{rev}\overline{p}(x) = p(x)$). Analogous simplifications occur for the T-even, $*$-even, and all the anti-variants in the scalar polynomial case.

Note that the strict alternation of matrix coefficients between symmetric and skew-symmetric (or Hermitian and skew-Hermitian) in $\star$-even/odd polynomials has also led to the use of the name *alternating* polynomial [65]. However, we will only use "alternating" as a collective term to denote any one of the six even or odd structures listed in Table 6.1.1. Similarly the word "palindromic" will typically be used as a generic term for any of the six structures in Table 6.1.1 that contain the word palindromic.

Two special matrices that play an important role in our investigation are the $k \times k$ reverse identity R_k in the context of palindromic structures, and the $k \times k$ diagonal matrix Σ_k of alternating signs in the context of even/odd structures:

$$R = R_k := \begin{bmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{bmatrix}_{k\times k} \quad \text{and} \quad \Sigma = \Sigma_k := \begin{bmatrix} (-1)^{k-1} & & 0 \\ & \ddots & \\ 0 & & (-1)^0 \end{bmatrix}_{k\times k} . \tag{6.1.4}$$

The subscript k will be dropped whenever it is clear from the context.

Remark 6.1.2. In Definition 6.1.1 the adjoint of an $n \times n$ matrix polynomial *could* have been defined with respect to the adjoint $\star$ of a more general scalar product (see [59]), rather than restricting $\star$ to just transpose or conjugate transpose. For example, a bilinear scalar product is defined by $\langle x, y\rangle := x^T My$ for some nonsingular matrix M, and adjoint $\star$ with respect to this scalar product is given by $A^\star = M^{-1}A^T M$. (Similarly a sesquilinear scalar product $\langle x, y\rangle := x^* My$ has adjoint given by $A^\star = M^{-1}A^* M$.) Then the definition of the corresponding matrix polynomial adjoint $P^\star(\lambda)$ would be formally identical to Definition 6.1.1, and the structures in Table 6.1.1 would make sense as written with $\star$ denoting the adjoint of a general scalar product. A well-known example of this more general notion is skew-Hamiltonian/Hamiltonian pencils [8], [64], which are $\star$-odd with respect to the symplectic form defined by $M = J = \left[\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right]$.

However, when the underlying scalar product is orthosymmetric [59] (concretely, if M satisfies $M^T = \varepsilon M$ for $\varepsilon = \pm 1$ in the bilinear case, or $M^* = \varepsilon M$, $|\varepsilon| = 1$, $\varepsilon \in \mathbb{C}$ in the sesquilinear case), then not much is gained by this apparent extra generality[1]. In the bilinear case we have

$$\begin{aligned} P(\lambda) \text{ is } \star\text{-palindromic} \quad &\Leftrightarrow \quad \operatorname{rev} P^\star(\lambda) = \operatorname{rev}\big(M^{-1}P^T(\lambda)M\big) = P(\lambda) \\ &\Leftrightarrow \quad \operatorname{rev}(MP(\lambda))^T = \operatorname{rev}\big(P^T(\lambda)M^T\big) = \varepsilon MP(\lambda)\,, \end{aligned}$$

so that $\star$-palindromicity of $P(\lambda)$ is equivalent to the T-(anti)-palindromicity of $MP(\lambda)$. Similar arguments show that $\star$-evenness or $\star$-oddness of $P(\lambda)$ is equivalent to the T-evenness or T-oddness of $MP(\lambda)$. Analogous results also hold for orthosymmetric sesquilinear forms, showing that in this case $\star$-structure reduces to $*$-structure. Thus for any of the standard scalar products with adjoint $\star$, the $\star$-structures in Table 6.1.1 can all be easily transformed into either the T or $*$ case. Note also that this reduction shows skew-Hamiltonian/Hamiltonian pencils to be equivalent to T-even or $*$-even pencils.

[1]Note that orthosymmetric scalar products include all the standard examples, which are either symmetric or skew-symmetric bilinear forms or Hermitian sesquilinear forms.

6.2 Spectral Symmetry

An important feature of the structured matrix polynomials in Table 6.1.1 is the special symmetry properties of their spectra, described in the following result. These eigenvalue pairings comprise the first of several ways in which alternating and palindromic matrix polynomials may reasonably be viewed as analogous to, and even as generalizations of Hamiltonian and symplectic matrices. Note that these eigenvalue pairings also extend to the more general $\star$-structures described in Remark 6.1.2.

Theorem 6.2.1 (Spectral Symmetry of Structured Matrix Polynomials). *Let $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$, $A_k \neq 0$ be a regular matrix polynomial that has any one of the palindromic or alternating structures listed in Table* 6.1.1*. Then the spectrum of $P(\lambda)$ has the pairing depicted in Table* 6.2.1*. Moreover, the algebraic, geometric, and partial multiplicities of the two eigenvalues in each such pair are equal.*
(Here $\lambda = 0$ is included as a possible eigenvalue, with $1/\lambda$ or $1/\overline{\lambda}$ to be interpreted as the eigenvalue ∞.)

Table 6.2.1: Spectral symmetries

Structure of $P(\lambda)$	eigenvalue pairing
(anti-)palindromic, T-(anti-)palindromic	$(\lambda, 1/\lambda)$
$*$-palindromic, $*$-anti-palindromic	$(\lambda, 1/\overline{\lambda})$
even, odd, T-even, T-odd	$(\lambda, -\lambda)$
$*$-even, $*$-odd	$(\lambda, -\overline{\lambda})$

Proof. We first recall some well-known facts [28], [30], [31] about the companion forms $C_1(\lambda)$ and $C_2(\lambda)$ of a regular matrix polynomial $P(\lambda)$:

- $P(\lambda)$ and $C_1(\lambda)$ have the same eigenvalues (including ∞) with the same algebraic, geometric, and partial multiplicities.
- $C_1(\lambda)$ and $C_2(\lambda)$ are always *strictly equivalent*, i.e., there exist nonsingular *constant* matrices E and F such that $C_1(\lambda) = E \cdot C_2(\lambda) \cdot F$.
- Strictly equivalent pencils have the same eigenvalues (including ∞), with the same algebraic, geometric, and partial multiplicities.

With these facts in hand, we first consider the case when $P(\lambda)$ is $\star$-palindromic or $\star$-anti-palindromic, so that $\operatorname{rev} P^\star(\lambda) = \chi_P P(\lambda)$ for $\chi_P = \pm 1$, equivalently $\chi_P A_i = A_{k-i}^\star$ for $i = 0:k$. Our strategy is to show that $C_1(\lambda)$ is strictly equivalent to $\operatorname{rev} C_1^\star(\lambda)$, from which the desired eigenvalue pairing and equality of multiplicities then follows. Using the nonsingular matrix

$$T := \begin{bmatrix} \chi_P I & & & 0 \\ & I & & \\ & & \ddots & \\ 0 & & & I \end{bmatrix} \cdot \begin{bmatrix} I & A_{k-1} & \cdots & A_1 \\ 0 & 0 & & -I \\ \vdots & & \iddots & \\ 0 & -I & & 0 \end{bmatrix},$$

we first show that $C_1(\lambda)$ is strictly equivalent to $\mathrm{rev}C_2^\star(\lambda)$.

$$\begin{aligned}
T\cdot C_1(\lambda)\cdot(R_k\otimes I_n) &= T\cdot\left(\lambda\begin{bmatrix}0 & & A_k\\ & \cdot^{\cdot^{\cdot}} \; I & \\ I & & 0\end{bmatrix}+\begin{bmatrix}A_0 & A_1 & \cdots & A_{k-1}\\ 0 & 0 & & -I\\ \vdots & & \cdot^{\cdot^{\cdot}} & \\ 0 & -I & & 0\end{bmatrix}\right)\\
&= \lambda\begin{bmatrix}\chi_P A_1 & \cdots & \chi_P A_{k-1} & \chi_P A_k\\ -I & & 0 & 0\\ & \ddots & & \vdots\\ 0 & & -I & 0\end{bmatrix}+\begin{bmatrix}\chi_P A_0 & & & 0\\ & I & & \\ & & \ddots & \\ 0 & & & I\end{bmatrix}\\
&= \lambda\begin{bmatrix}A_{k-1} & -I & & 0\\ \vdots & & \ddots & \\ A_1 & 0 & & -I\\ A_0 & 0 & \cdots & 0\end{bmatrix}^\star+\begin{bmatrix}A_k & & & 0\\ & I & & \\ & & \ddots & \\ 0 & & & I\end{bmatrix}^\star\\
&= \mathrm{rev}C_2^\star(\lambda)\,.
\end{aligned}$$

But $\mathrm{rev}C_2^\star(\lambda)$ is always strictly equivalent to $\mathrm{rev}C_1^\star(\lambda)$, since $C_1(\lambda)$ and $C_2(\lambda)$ are. This completes the proof for this case.

For the case of "purely" palindromic or anti-palindromic matrix polynomials, i.e., polynomials $P(\lambda)$ satisfying $\mathrm{rev}P(\lambda)=\chi_P P(\lambda)$, an analogous computation shows that

$$T\cdot C_1(\lambda)\cdot(R_k\otimes I_n)=\mathrm{rev}C_1(\lambda)\,.$$

Thus $C_1(\lambda)$ is strictly equivalent to $\mathrm{rev}C_1(\lambda)$, which again implies the desired eigenvalue pairing and equality of multiplicities.

Next assume that $P(\lambda)$ is $\star$-even or $\star$-odd, so $P^\star(-\lambda)=\varepsilon_P P(\lambda)$ for $\varepsilon_P=\pm1$. We show that $C_1(\lambda)$ is strictly equivalent to $C_1^\star(-\lambda)$, from which the desired pairing of eigenvalues and equality of multiplicities follows. The following calculation shows that $C_1(\lambda)$ is strictly equivalent to $C_2^\star(-\lambda)$:

$$\begin{aligned}
&\big(\mathrm{diag}(\varepsilon_P,-\Sigma_{k-1})\otimes I_n\big)\cdot C_1(\lambda)\cdot(\Sigma_k\otimes I_n)\\
&= \lambda\begin{bmatrix}\varepsilon_P(-1)^{k-1}A_k & & & 0\\ & -I & & \\ & & \ddots & \\ 0 & & & -I\end{bmatrix}+\begin{bmatrix}\varepsilon_P(-1)^{k-1}A_{k-1} & \cdots & \varepsilon_P(-1)^1A_1 & \varepsilon_P A_0\\ -I & & 0 & 0\\ & \ddots & & \vdots\\ 0 & & -I & 0\end{bmatrix}\\
&= -\lambda\begin{bmatrix}A_k & & & 0\\ & I & & \\ & & \ddots & \\ 0 & & & I\end{bmatrix}^\star+\begin{bmatrix}A_{k-1} & -I & & 0\\ \vdots & & \ddots & \\ A_1 & 0 & & -I\\ A_0 & 0 & \cdots & 0\end{bmatrix}^\star = C_2^\star(-\lambda)\,.
\end{aligned}$$

The strict equivalence of $C_2^\star(-\lambda)$ and $C_1^\star(-\lambda)$ now follows from that of $C_2(\lambda)$ and $C_1(\lambda)$, and the proof for this case is complete.

For "purely" even or odd polynomials $P(\lambda)$, that is $P(-\lambda)=\varepsilon_P P(\lambda)$, an analogous computation

$$\big(\mathrm{diag}(\varepsilon_P,-\Sigma_{k-1})\otimes I_n\big)\cdot C_1(\lambda)\cdot(\Sigma_k\otimes I_n)=C_1(-\lambda)$$

shows that $C_1(\lambda)$ is strictly equivalent to $C_1(-\lambda)$, which implies the desired eigenvalue pairing and equality of multiplicities. □

If the coefficient matrices of P are real, then the eigenvalues of a $\star$-even or $\star$-odd matrix polynomial occur in quadruples $(\lambda, \overline{\lambda}, -\lambda, -\overline{\lambda})$. This property has sometimes been referred to as "Hamiltonian spectral symmetry", since the eigenvalues of real Hamiltonian matrices have such symmetry [61], [65]. Note however that this is actually a feature common to matrices in Lie algebras associated with any real scalar product, and is not confined to Hamiltonian matrices [59]. Similarly, the eigenvalues of real $\star$-palindromic and anti-$\star$-palindromic matrix polynomials occur not just in pairs but in quadruples $(\lambda, \overline{\lambda}, 1/\lambda, 1/\overline{\lambda})$, a property sometimes referred to as "symplectic spectral symmetry", since real symplectic matrices exhibit this behavior. But once again, this type of eigenvalue symmetry is an instance of a more general phenomenon associated with matrices in the Lie group of any real scalar product, such as the real pseudo-orthogonal (Lorentz) groups. See [24], [50], [61] for detailed coverage of Hamiltonian and symplectic matrices, and [32], [59] for properties of matrices in the Lie algebra or Lie group of more general scalar products.

Remark 6.2.2. The ability to directly convert a structured matrix into an equivalent structured pencil is a second aspect of the analogy between Hamiltonian/symplectic matrices and alternating/palindromic polynomials. To any matrix A one naturally associates the pencil $\lambda I - A$, and so for Hamiltonian H consider the pencil $\lambda I - H$. Then equivalent to $\lambda I - H$ is the T-even pencil $J(\lambda I - H) = \lambda J - JH$. This simple conversion provides a very direct sense in which alternating polynomials generalize Hamiltonian matrices. Indeed, one can convert a matrix from any Lie algebra associated with an orthosymmetric scalar product into either an even or odd pencil by a similar procedure, thus showing that alternating polynomials generalize an even broader range of well-known structured matrices.

The situation for symplectic matrices is much less straightforward. The conversion of pencils $\lambda I - S$ with a symplectic S to some type of palindromic pencil certainly cannot be uniformly achieved with a fixed equivalence like the one used above for Hamiltonian matrices. However, some recent work of Schröder [75] shows that almost all symplectic matrices S have a factorization of the form $S = Z^{-1}Z^T$ for some nonsingular Z, and hence almost all pencils $\lambda I - S$ with a symplectic S are equivalent to a T-anti-palindromic pencil $\lambda Z - Z^T$. The only symplectic S that do not admit such a factorization are those having an odd number of even-sized Jordan blocks for the eigenvalue $+1$.

What about converting $\lambda I - S$ into a T-palindromic pencil $\lambda W + W^T$? As an immediate corollary of the results in [75] one sees that the only obstruction to doing this is the Jordan structure of S at the eigenvalue -1; a pencil $\lambda I - S$ with a symplectic S is equivalent to a T-palindromic pencil if and only if S has an even number of even-sized Jordan blocks for the eigenvalue -1.

Several natural questions, though, remain open. If, instead of looking at only pencils, one considers "representations" of symplectic matrices by palindromic polynomials of degree $k \geq 2$, can *every* symplectic matrix now be brought under the palindromic umbrella? In a different direction, one might ask whether Schröder's results relating symplectic matrices to palindromic pencils extend to structured matrices from other Lie groups associated with scalar products.

6.3 Cayley Transformations

It is well known that the *Cayley transformation* of matrices [50], [69, p. 103–105] and its generalization to pencils [50], [62] relates Hamiltonian to symplectic structure for both matrices and pencils. By extending the classical definition of this transformation to matrix polynomials, we now develop analogous relationships between palindromic and alternating polynomials. These relationships constitute yet a third aspect of the analogy between Hamiltonian/symplectic matrices and alternating/palindromic polynomials.

Our choice of definition is motivated by the following observation: the only Möbius transformations of the complex plane that map reciprocal pairs $(\mu, 1/\mu)$ to plus/minus pairs $(\lambda, -\lambda)$ are $\alpha\big(\frac{\mu-1}{\mu+1}\big)$ and $\beta\big(\frac{1+\mu}{1-\mu}\big)$, where $\alpha,\ \beta \in \mathbb{C}$ are nonzero constants. When $\alpha = \beta = 1$, these transformations also map conjugate reciprocal pairs $(\mu, 1/\overline{\mu})$ to conjugate plus/minus pairs $(\lambda, -\overline{\lambda})$. Putting this together with Theorem 6.2.1, we see that the Möbius transformations $\frac{\mu-1}{\mu+1}$ and $\frac{1+\mu}{1-\mu}$ translate the spectral symmetries of palindromic polynomials to those of alternating matrix polynomials. Consequently, it is reasonable to anticipate that Cayley transformations modelled on these particular Möbius transformations might have an analogous effect on structure at the level of matrix polynomials. These observations therefore lead us to adopt the following definition as the natural extension (in this context) of the Cayley transformation to matrix polynomials.

Definition 6.3.1 (Cayley Transformations of Matrix Polynomials). Let $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ be a matrix polynomial of degree k. Then the matrix polynomials

$$\mathcal{C}_{-1}(P)(\mu) := (\mu+1)^k P\left(\frac{\mu-1}{\mu+1}\right) \quad \text{and} \quad \mathcal{C}_{+1}(P)(\mu) := (1-\mu)^k P\left(\frac{1+\mu}{1-\mu}\right) \tag{6.3.1}$$

are the *Cayley transformations of* $P(\lambda)$ *with pole at* -1 *or* $+1$, respectively.

When viewed as maps on the space of $n \times n$ matrix polynomials of degree $k \geq 1$, the Cayley transformations in (6.3.1) can be shown by a direct calculation to be inverses of each other, up to a scaling factor.

Proposition 6.3.2. *For any $n \times n$ matrix polynomial P of degree $k \geq 1$ we have $\mathcal{C}_{+1}(\mathcal{C}_{-1}(P)) = \mathcal{C}_{-1}(\mathcal{C}_{+1}(P)) = 2^k \cdot P$.*

The next lemma gives some straightforward observations about how the adjoint and reversal operations from Definition 6.1.1 interact with the Cayley transformations $\mathcal{C}_{-1}$ and $\mathcal{C}_{+1}$. These will be helpful in showing how structure in a matrix polynomial leads to structure in its Cayley transformations.

Lemma 6.3.3. *Let P be a matrix polynomial of degree $k \geq 1$. Then*

$$\big(\mathcal{C}_{-1}(P)\big)^{\star}(\mu) = \mathcal{C}_{-1}(P^{\star})(\mu), \qquad \big(\mathcal{C}_{+1}(P)\big)^{\star}(\mu) = \mathcal{C}_{+1}(P^{\star})(\mu), \tag{6.3.2}$$

$$\operatorname{rev}\big(\mathcal{C}_{-1}(P)\big)^{\star}(\mu) = (\mu+1)^k P^{\star}\left(-\frac{\mu-1}{\mu+1}\right), \quad \mu \neq -1, \tag{6.3.3a}$$

$$\operatorname{rev}\big(\mathcal{C}_{+1}(P)\big)^{\star}(\mu) = (-1)^k(1-\mu)^k P^{\star}\left(-\frac{1+\mu}{1-\mu}\right), \quad \mu \neq 1. \tag{6.3.3b}$$

Proof. The proof of (6.3.2) is straightforward. We only prove (6.3.3b); the proof of (6.3.3a) is similar. Since $\mathcal{C}_{+1}(P)$ and hence $\mathcal{C}_{+1}(P)^\star$ are matrix polynomials of degree k,

$$\begin{aligned}
\operatorname{rev}\big(\mathcal{C}_{+1}(P)\big)^\star(\mu) &= \mu^k\big(\mathcal{C}_{+1}(P)\big)^\star\left(\frac{1}{\mu}\right) = \mu^k\mathcal{C}_{+1}(P^\star)\left(\frac{1}{\mu}\right) && \text{by (6.3.2), (6.1.3)}\\
&= \mu^k(1-1/\mu)^kP^\star\left(\frac{1+1/\mu}{1-1/\mu}\right) && \text{by (6.3.1)}\\
&= (-1)^k(1-\mu)^kP^\star\left(-\frac{1+\mu}{1-\mu}\right). \quad \square
\end{aligned}$$

We now gather together in Table 6.3.1 all the details of the relationships between palindromic/alternating structure in polynomials and in their Cayley transforms. Note that these relationships sometimes depend on the parity of the degree of the given polynomial.

Theorem 6.3.4 (Structure Correspondence via Cayley).
Let $P(\lambda)$ be a matrix polynomial of degree $k \geq 1$. Then the correspondence between structure in P and in its Cayley transforms $\mathcal{C}_{-1}(P)$ and $\mathcal{C}_{+1}(P)$ is as stated in Table 6.3.1*. (Note that each structure correspondence in this table is an if and only if statement.)*

Table 6.3.1: Cayley transformations of structured matrix polynomials

<table>
<tr><th rowspan="2">$P(\lambda)$</th><th colspan="2">$\mathcal{C}_{-1}(P)(\mu)$</th><th colspan="2">$\mathcal{C}_{+1}(P)(\mu)$</th></tr>
<tr><th>k even</th><th>k odd</th><th>k even</th><th>k odd</th></tr>
<tr><td>palindromic
$\star$-palindromic</td><td>even
$\star$-even</td><td>odd
$\star$-odd</td><td colspan="2">even
$\star$-even</td></tr>
<tr><td>anti-palindromic
$\star$-anti-palindromic</td><td>odd
$\star$-odd</td><td>even
$\star$-even</td><td colspan="2">odd
$\star$-odd</td></tr>
<tr><td>even
$\star$-even</td><td colspan="2">palindromic
$\star$-palindromic</td><td>palindromic
$\star$-palindromic</td><td>anti-palindromic
$\star$-anti-palindromic</td></tr>
<tr><td>odd
$\star$-odd</td><td colspan="2">anti-palindromic
$\star$-anti-palindromic</td><td>anti-palindromic
$\star$-anti-palindromic</td><td>palindromic
$\star$-palindromic</td></tr>
</table>

Proof. Since the proofs of these structure correspondences are all similar, only one of them is given here. We show that $P(\lambda)$ is $\star$-even if and only if $\mathcal{C}_{+1}(P)(\mu)$ is $\star$-palindromic when k is even and $\star$-anti-palindromic when k is odd. Now $P(\lambda)$ being $\star$-even is equivalent, by definition, to $P^\star(-\lambda) = P(\lambda)$ for all λ. Setting $\lambda = \frac{1+\mu}{1-\mu}$ and multiplying by $(1-\mu)^k$ yields

$$\begin{aligned}
P(\lambda) \text{ is } \star\text{-even} &\iff (1-\mu)^kP^\star\left(-\frac{1+\mu}{1-\mu}\right) = (1-\mu)^kP\left(\frac{1+\mu}{1-\mu}\right) \quad \text{for all } \mu \neq 1\\
&\iff (-1)^k\operatorname{rev}(\mathcal{C}_{+1}(P))^\star(\mu) = \mathcal{C}_{+1}(P)(\mu) \quad \text{by Lemma 6.3.3,}
\end{aligned}$$

from which the desired result follows. $\square$

Observe that the results in Table 6.3.1 are consistent with $\mathcal{C}_{-1}(P)$ and $\mathcal{C}_{+1}(P)$ being essentially inverses of each other, as shown in Proposition 6.3.2.

6.4 Applications

To illustrate the practical importance of palindromic and alternating matrix polynomials, we conclude this chapter with a sampling of applications that lead to polynomial eigenvalue problems with one of these structures. Many of these problems arise in the analysis and numerical solution of higher order systems of ordinary and partial differential equations, as in the first two examples.

Example 6.4.1. (Quadratic complex T-palindromic matrix polynomials) A project of the company SFE GmbH in Berlin investigates rail traffic noise caused by high speed trains [37], [38]. The vibration of an infinite rail track is simulated and analyzed to obtain information on the development of noise between wheel and rail. In the model, the rail is assumed to be infinite and is tied to the ground on sleepers, where neighboring sleepers are spaced $s = 0.6$ m apart (including the width of one of the sleepers). This segment of the infinite track is called a sleeper bay. The part of the rail corresponding to one sleeper bay is then discretized using classical finite element methods for the model of excited vibration (Figure 6.4.1).

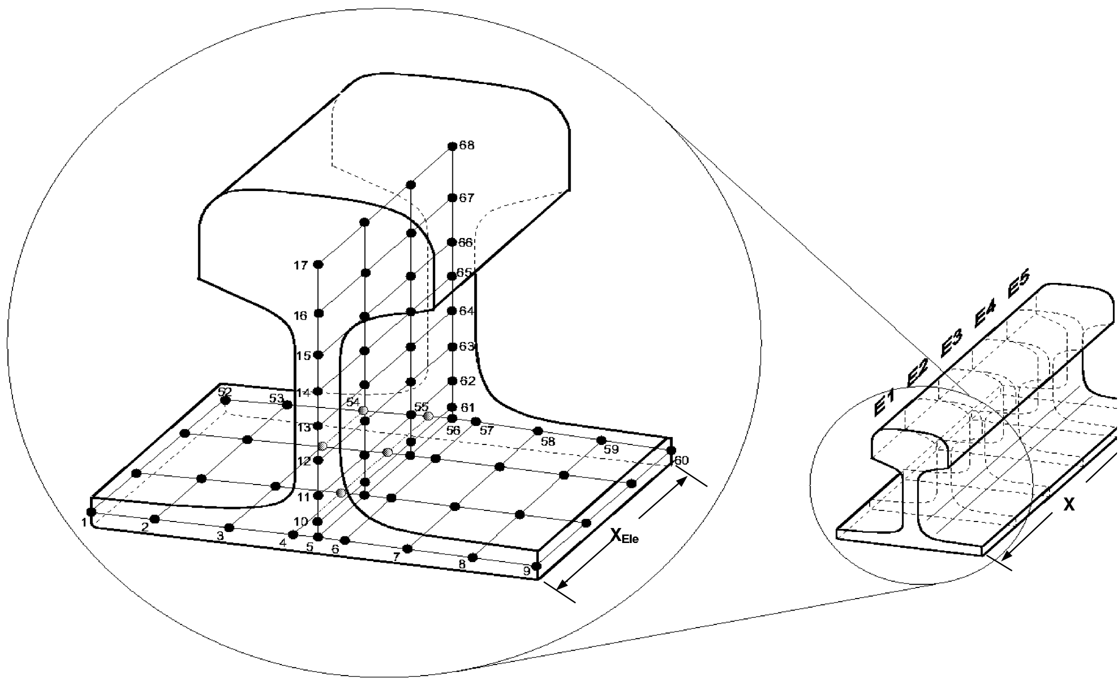


Figure 6.4.1: FE discretization of the rail in one sleeper bay.

The discretization leads to an infinite dimensional second order system of the form $M\ddot{x} + D\dot{x} + Sx = F$, with infinite block-tridiagonal real symmetric coefficient matrices M, D, S, where

$$M = \begin{bmatrix} \ddots & \ddots & 0 & \dots & 0 \\ \ddots & M_{j-1,0} & M_{j,1} & 0 & \vdots \\ 0 & M_{j,1}^T & M_{j,0} & M_{j+1,1} & 0 \\ \vdots & \ddots & M_{j+1,1}^T & M_{j+1,0} & \ddots \\ 0 & \dots & 0 & \ddots & \ddots \end{bmatrix}, \; x = \begin{bmatrix} \vdots \\ x_{j-1} \\ x_j \\ x_{j+1} \\ \vdots \end{bmatrix}, \; F = \begin{bmatrix} \vdots \\ F_{j-1} \\ F_j \\ F_{j+1} \\ \vdots \end{bmatrix},$$

and where D, S have the same block structure as M with blocks $D_{j,0}, D_{j,1}$ and $S_{j,0}, S_{j,1}$, respectively. Here, $M_{j,0}$ is symmetric positive definite and $D_{j,0}, S_{j,0}$ are symmetric positive semidefinite for all j.

There are several ways to approach the solution of the problem, which presents a mixture between a differential equation (time derivatives of x) and a difference equation (space differences in j).

Since one is interested in studying the behavior of the system under excitation, one makes the ansatz $F_j = \hat{F}_j e^{i\omega t}$, $x_j = \hat{x}_j e^{i\omega t}$, where ω is the excitation frequency. This leads to a second order difference equation with variable coefficients for the $\hat{x}_j$ given by

$$A_{j-1,j}^T \hat{x}_{j-1} + A_{jj}\hat{x}_j + A_{j,j+1}\hat{x}_{j+1} = \hat{F}_j,$$

with the coefficient matrices

$$A_{j,j+1} = -\omega^2 M_{j,1} + i\omega D_{j,1} + K_{j,1}, \qquad A_{jj} = -\omega^2 M_{j,0} + i\omega D_{j,0} + K_{j,0}.$$

Observing that the system matrices vary periodically due to the identical form of the rail track in each sleeper bay, we may combine the (say ℓ) parts belonging to the rail in one sleeper bay into one vector

$$y_j = \begin{bmatrix} \hat{x}_j \\ \hat{x}_{j+1} \\ \vdots \\ \hat{x}_{j+\ell} \end{bmatrix},$$

and thus obtain a constant coefficient second order difference equation

$$A_1^T y_{j-1} + A_0 y_j + A_1 y_{j+1} = G_j$$

with coefficient matrices

$$A_0 = \begin{bmatrix} A_{j,j} & A_{j,j+1} & & 0 \\ A_{j,j+1}^T & A_{j+1,j+1} & \ddots & \\ & \ddots & \ddots & A_{j+\ell-1,j+l} \\ 0 & & A_{j+\ell-1,j+\ell}^T & A_{j+\ell,j+\ell} \end{bmatrix}, \; A_1 = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 \\ A_{j+\ell,j+\ell+1} & 0 & \dots & 0 \end{bmatrix},$$

that depend on the frequency ω. For this system we then make the ansatz $y_{j+1} = \kappa y_j$, which leads to the complex eigenvalue problem

$$\frac{1}{\kappa}(A_1^T + \kappa A_0 + \kappa^2 A_1) y = 0.$$

Clearly, the underlying matrix polynomial $A_1^T + \kappa A_0 + \kappa^2 A_1$ is T-palindromic, because A_0 is complex symmetric, i.e., $A_0 = A_0^T$. It should be noted that in this application A_1 is highly rank-deficient.

Example 6.4.2. (Quadratic real and complex T-palindromic polynomials)
In [89] the mathematical modelling and numerical simulation of the behavior of periodic surface acoustic wave (SAW) filters is discussed. SAW-filters are piezoelectric devices used in telecommunications, e.g., TV-sets and cell phones, for frequency filtering; other kinds of SAW-devices find application in radar and sensor technology as well as in the field of non-destructive evaluation. In modelling these devices, Floquet-Bloch theory is used in [89] to replace the underlying periodic structure of the problem by a single reference cell together with quasi-periodic boundary conditions. This Bloch-ansatz reduces the problem to calculating the so-called "dispersion diagram", i.e., the functional relation between the excitation frequency ω and the (complex) propagation constant γ. A finite element discretization then leads to a parameter-dependent Galerkin system, which upon further reduction (and invocation of the quasi-periodic boundary conditions) becomes a T-palindromic quadratic eigenvalue problem

$$(\gamma^2 A + \gamma B + A^T)v = 0, \quad \text{with} \quad B^T = B.$$

Note that A and B both depend on the parameter ω. If absorbing boundary conditions (necessary for volume wave radiation) are included in the model, then A and B are complex, otherwise real.

Example 6.4.3. (Quadratic $*$-palindromic matrix polynomials)
In [36], bisection and level set methods are presented to compute the Crawford number

$$\gamma(A,B) := \min_{\substack{z \in \mathbb{C}^n \\ \|z\|_2 = 1}} \sqrt{(z^*Az)^2 + (z^*Bz)^2}$$

for two Hermitian matrices $A, B \in \mathbb{C}^{n\times n}$. It is shown in [36, Theorem 2.2] that $\gamma(A,B)$ measures the distance of a Hermitian pair (A,B) to the nearest non-definite pair in the 2-norm. From [36, formula (2.8)]

$$\gamma(A,B) = \max\left(\max_{0\le\theta\le 2\pi} \lambda_{\min}(A\cos\theta + B\sin\theta), 0\right),$$

the problem of computing the Crawford number can be reduced to the computation of

$$\max\left\{\lambda_{\min}(M(z)) \,:\, |z| = 1\right\},$$

where $M(z) = (z^{-1}C + zC^*)/2$ and $C = A + iB$. It is easy to check that $M(z)$ is Hermitian for any z on the unit circle. Since for a given $\xi \in \mathbb{R}$, the equivalence

$$\det\big(M(z) - \xi I\big) = 0 \iff \det\big(C - 2\xi z I + z^2 C^*\big) = 0$$

holds, the authors of [36] discuss the following strategy as a base for a bisection algorithm. Select a value $\xi \in \mathbb{R}$ and compute the $2n$ eigenvalues z_j of the $*$-palindromic matrix polynomial $P(z) = C - 2\xi z I + z^2 C^*$. For each z_j on the unit circle compute the smallest eigenvalue $\lambda_{\min}(M(z_j))$ of $M(z_j)$. If $\lambda_{\min}(M(z_j)) = \xi$ then $\gamma(A,B) \ge \lambda_{\min}(M(z_j))$; otherwise we have $\gamma(A,B) < \lambda_{\min}(M(z_j))$. Thus $\gamma(A,B)$ can be approximated via a bisection method.

Example 6.4.4. (Quadratic T-even matrix polynomials)
The study of corner singularities in anisotropic elastic materials [4], [5], [54], [64] leads to quadratic eigenvalue problems of the form

$$P(\lambda)v = (\lambda^2 M + \lambda G + K)v = 0,$$

with $M = M^T$, $G = -G^T$, $K = K^T$ in $\mathbb{R}^{n\times n}$. The coefficient matrices are large and sparse, having been produced by a finite element discretization. Here, M is a positive definite mass matrix and $-K$ is a stiffness matrix. Since the coefficient matrices alternate between real symmetric and skew-symmetric matrices we see that $P^T(-\lambda) = P(\lambda)$, and thus the matrix polynomial is T-even.

Gyroscopic systems [47], [83] also lead to quadratic T-even matrix polynomials.

Example 6.4.5. (Higher degree $*$-even matrix polynomials)
The linear quadratic optimal control problem for higher order systems of ordinary differential equations leads to the two-point boundary value problem for $2(k-1)$th order ordinary differential equations of the form

$$\sum_{j=1}^{k-1} \begin{bmatrix} (-1)^{j-1}Q_j & M_{2j}^* \\ M_{2j} & 0 \end{bmatrix} \begin{bmatrix} x^{(2j)} \\ \mu^{(2j)} \end{bmatrix} + \sum_{j=1}^{k-1} \begin{bmatrix} 0 & -M_{2j-1}^* \\ M_{2j-1} & 0 \end{bmatrix} \begin{bmatrix} x^{(2j-1)} \\ \mu^{(2j-1)} \end{bmatrix} + \begin{bmatrix} -Q_0 & M_0^* \\ M_0 & -BW^{-1}B^* \end{bmatrix} \begin{bmatrix} x \\ \mu \end{bmatrix} = 0\,,$$

where W and Q_j are Hermitian for $j = 1, \ldots, k-1$, see [4], [5], [61], [65]. The substitution $\left[\begin{smallmatrix} x \\ \mu \end{smallmatrix}\right] = e^{\lambda t}v$ then yields the eigenvalue problem $P(\lambda)v = 0$ with the underlying $*$-even matrix polynomial of degree $2(k-1)$ given by

$$\sum_{j=1}^{k-1} \left(\lambda^{2j} \begin{bmatrix} (-1)^{j-1}Q_j & M_{2j}^* \\ M_{2j} & 0 \end{bmatrix} + \lambda^{2j-1} \begin{bmatrix} 0 & -M_{2j-1}^* \\ M_{2j-1} & 0 \end{bmatrix} \right) + \begin{bmatrix} -Q_0 & M_0^* \\ M_0 & -BW^{-1}B^* \end{bmatrix}.$$

Example 6.4.6. (Higher degree $*$-palindromic matrix polynomials)
Consider the discrete time optimal control problem to minimize

$$\sum_{j=0}^{\infty} \begin{bmatrix} x_j \\ u_j \end{bmatrix}^* \mathcal{H} \begin{bmatrix} x_j \\ u_j \end{bmatrix}, \qquad \mathcal{H} = \begin{bmatrix} Q & Y \\ Y^* & R \end{bmatrix} \tag{6.4.1}$$

subject to the discrete time control

$$\sum_{i=0}^{k} M_i x_{j+k-i} = Bu_j, \tag{6.4.2}$$

with initial conditions $x_0, x_1, \ldots, x_{k-1}$ given. Here the matrices are of size $Q, M_i \in \mathbb{F}^{n\times n}$ for $i = 0, \ldots, k$, $R \in \mathbb{F}^{m\times m}$, and $Y, B \in \mathbb{F}^{n\times m}$, and satisfy $Q^* = Q$, $R^* = R$. (We discuss only the even degree case $k = 2\ell$; the odd degree case is similar but notationally more involved.)

In the classical application from linear quadratic optimal control, the matrix $\mathcal{H}$ in (6.4.1) is symmetric or Hermitian positive semidefinite, with R being positive definite. In applications from discrete time H_∞ control, however, both matrices may be indefinite and singular.

The standard way to attack this problem is to turn it into a first order system and then apply well-known techniques for such systems (see, e.g., [61]), leading to a two-point boundary value problem whose solution can be found by solving a generalized eigenvalue problem for a $(2kn + m) \times (2kn + m)$ pencil of the form

$$\mathcal{L}(\lambda) = \lambda \begin{bmatrix} 0 & \mathcal{E} & 0 \\ \mathcal{A}^* & 0 & 0 \\ \mathcal{B}^* & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & \mathcal{A} & \mathcal{B} \\ \mathcal{E}^* & \mathcal{G} & \mathcal{Y} \\ 0 & \mathcal{Y}^* & \mathcal{R} \end{bmatrix}. \tag{6.4.3}$$

For a large class of optimal control problems there is sufficient special structure within the blocks of $\mathcal{L}(\lambda)$ that it can be reduced to a *symplectic pencil*, and sometimes even further all the way to a standard eigenvalue problem for a symplectic *matrix* [61]. However, for many applications such reductions are not possible.

On the other hand, in all cases it is always possible to undo the conversion to first order that produced (6.4.3), and to do it in such a way as to obtain a degree k polynomial eigenvalue problem for the $(2n + m) \times (2n + m)$ matrix polynomial

$$P_s(\lambda) = \lambda^{2\ell} \begin{bmatrix} 0 & M_0 & 0 \\ M_{2\ell}^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \lambda^{2\ell-1} \begin{bmatrix} 0 & M_1 & 0 \\ M_{2\ell-1}^* & Q & 0 \\ 0 & Y^* & 0 \end{bmatrix} + \lambda^{2\ell-2} \begin{bmatrix} 0 & M_2 & 0 \\ M_{2\ell-2}^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
$$+ \cdots + \lambda^2 \begin{bmatrix} 0 & M_{2\ell-2} & 0 \\ M_2^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 & M_{2\ell-1} & 0 \\ M_1^* & 0 & 0 \\ -B^* & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & M_{2\ell} & -B \\ M_0^* & 0 & Y \\ 0 & 0 & R \end{bmatrix}.$$

The polynomial $P_s(\lambda)$ is not palindromic, but by using a non-equivalence transformation in a manner analogous to a technique used in [88], $P_s(\lambda)$ can be very simply transformed into a palindromic polynomial. Indeed, multiplying $P_s(\lambda)$ on the left by $\operatorname{diag}(\lambda^{\ell-1} I_n, I_n, \lambda^\ell I_m)$ and on the right by $\operatorname{diag}(I_n, \lambda^{1-\ell} I_n, I_m)$ leads to the degree $k = 2\ell$ $*$-palindromic matrix polynomial

$$P_p(\lambda) = \lambda^{2\ell} \begin{bmatrix} 0 & M_0 & 0 \\ M_{2\ell}^* & 0 & 0 \\ 0 & Y^* & 0 \end{bmatrix} + \lambda^{2\ell-1} \begin{bmatrix} 0 & M_1 & 0 \\ M_{2\ell-1}^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \cdots + \lambda^{\ell+2} \begin{bmatrix} 0 & M_{\ell-2} & 0 \\ M_{\ell+2}^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
$$+ \lambda^{\ell+1} \begin{bmatrix} 0 & M_{\ell-1} & 0 \\ M_{\ell+1}^* & 0 & 0 \\ -B^* & 0 & 0 \end{bmatrix} + \lambda^{\ell} \begin{bmatrix} 0 & M_\ell & 0 \\ M_\ell^* & Q & 0 \\ 0 & 0 & R \end{bmatrix} + \lambda^{\ell-1} \begin{bmatrix} 0 & M_{\ell+1} & -B \\ M_{\ell-1}^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
$$+ \lambda^{\ell-2} \begin{bmatrix} 0 & M_{\ell+2} & 0 \\ M_{\ell-2}^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \cdots + \lambda \begin{bmatrix} 0 & M_{2\ell-1} & 0 \\ M_1^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & M_{2\ell} & 0 \\ M_0^* & 0 & Y \\ 0 & 0 & 0 \end{bmatrix}.$$

Since $\det P_p(\lambda) = \lambda^{\ell m} \det P_s(\lambda)$, it follows that $P_s(\lambda)$ and $P_p(\lambda)$ have the same finite eigenvalues (counted with multiplicities) except for ℓm additional zero eigenvalues of $P_p(\lambda)$.

There is an alternative way of formulating the discrete time optimal control problem as a $*$-palindromic polynomial. Picking up the story with the pencil $\mathcal{L}(\lambda)$ in (6.4.3), first make the change of variable $\lambda = -\mu^2$ to get the quadratic polynomial

$$\mathcal{Q}(\mu) \;:=\; -\mathcal{L}(-\mu^2) \;=\; \mu^2 \begin{bmatrix} 0 & \mathcal{E} & 0 \\ \mathcal{A}^* & 0 & 0 \\ \mathcal{B}^* & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \mathcal{A} & \mathcal{B} \\ \mathcal{E}^* & \mathcal{G} & \mathcal{Y} \\ 0 & \mathcal{Y}^* & \mathcal{R} \end{bmatrix}. \tag{6.4.4}$$

The polynomial $\mathcal{Q}(\mu)$ is still not $*$-palindromic, but a non-equivalence transformation like the one used above to convert $P_s(\lambda)$ into $P_p(\lambda)$ will make it into one. Specifically, premultiply $\mathcal{Q}(\mu)$ by $\mathrm{diag}(I_{kn}\,, \mu I_{kn}\,, \mu I_m)$ and postmultiply by $\mathrm{diag}\big((\mu)^{-1} I_{kn}\,, I_{kn}\,, I_m\big)$ to get the polynomial

$$\mathcal{Q}_p(\mu) \;:=\; \mu^2 \begin{bmatrix} 0 & \mathcal{E} & 0 \\ \mathcal{A}^* & 0 & 0 \\ \mathcal{B}^* & 0 & 0 \end{bmatrix} + \mu \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathcal{G} & \mathcal{Y} \\ 0 & \mathcal{Y}^* & \mathcal{R} \end{bmatrix} + \begin{bmatrix} 0 & \mathcal{A} & \mathcal{B} \\ \mathcal{E}^* & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{6.4.5}$$

Since the blocks $\mathcal{G}$ and $\mathcal{R}$ are Hermitian, the middle term of $\mathcal{Q}_p(\mu)$ is Hermitian, and so $\mathcal{Q}_p(\mu)$ is $*$-palindromic. Since $\det \mathcal{Q}_p(\mu) = \mu^m \det \mathcal{Q}(\mu)$, it follows that $\mathcal{Q}_p(\mu)$ and $\mathcal{Q}(\mu)$ have the same finite eigenvalues (counted with multiplicities) except for m additional zero eigenvalues of $\mathcal{Q}_p(\mu)$.

Remark 6.4.7. Example 6.4.6 illustrates yet another sense in which it is reasonable to view palindromic and alternating polynomials as generalizations of symplectic and Hamiltonian matrices. It is not just the sharing of mathematical properties like spectral symmetries and parallel relations via Cayley transformation that gives this analogy some substance. It is also by the extension of the range of applicability to a wider class of practical problems that palindromic polynomials can be seen to generalize symplectic structure.

Remark 6.4.8. The alert reader may have noticed the *absence* of "purely" palindromic or "purely" even/odd polynomials, that is to say ones not involving the adjoint $\star$ in their definition, among the applications described in this section. As of this writing we are not aware of any application that leads to any of these "pure" structures. For this reason (and others to be discussed in the next chapter), these four types of structured polynomial will not be considered very much in the further development of this topic . The focus instead will be on the eight $\star$-structures defined in Table 6.1.1.

Remark 6.4.9. Finally, it should be noted that palindromic structure has a (potentially) very important computational advantage over symplectic structure. Palindromicity is a *linear* structure, and so is (or at least should be) numerically easier to maintain in the face of rounding errors than a *nonlinear* structure like symplecticity (either for matrices or pencils). (Note that the defining conditions for symplectic matrices and for symplectic pencils put *quadratic constraints* on the matrix entries.) Thus the process of replacing a symplectic matrix or pencil by a palindromic polynomial can already in some sense be viewed as a kind of "linearizing" of the problem.

6.5 A Palindrome Sampler

The palindrome is a form of wordplay that appears in many languages and has an ancient history. The earliest palindromes have been attributed to Sotades of Maroneia in Thrace during the third century B.C.; thus palindromes have sometimes been referred to as "Sotadics" or "Sotadic verses" [19]. Just for fun, here's a selection of palindromes in various languages, with pointers to palindromes in other media and other forms. For further information (and larger collections of palindromes than you could ever want) see the website `www.nyx.net/~jkalb/palindromes/` with its many useful links, as well as [9], [13], [15], [19], [20], [21], [53], and [66].

The earliest recorded palindrome in English dates from 1614 — "Lewd did I live & evil I did dwel" — and is attributed to John Taylor, the self-proclaimed "Water Poet" of London. Here are some more palindromes in English, several with a mathematical flavor:

Mom
Rotator
I prefer pi.
Satire: Veritas.
Rise to vote, sir.
So many dynamos!
Never odd or even.
Niagara, O roar again!
Won't lovers revolt now?
Able was I ere I saw Elba.
A man, a plan, a canal – Panama.
Satan, oscillate my metallic sonatas.
I, man, am regal — a German am I.
Go hang a salami! I'm a lasagna hog!
Sums are not set as a test on Erasmus.
Anne, I vote more cars race Rome-to-Vienna.
Barclay ordered an omelette, lemonade, red royal crab.
Are we not drawn onward, we few, drawn onward to new era?

Doc, note, I dissent. A fast never prevents a fatness. I diet on cod.
— *Peter Hilton, 1947* [39] [53, p. 287]

T. Eliot, top bard, notes putrid tang emanating, is sad. I'd assign it a name: "gnat dirt". Upset on drab pot toilet.
— *Alastair Reid, 1959*

Although English is certainly replete with palindromes, there are many examples in other languages as well. Indeed, Finnish has sometimes been termed the "language of palindromes", perhaps because it contains a relatively large number of single-word palindromes. The Finnish word for soap salesman, "saippuakauppias", has been claimed by some as the longest single-word palindrome in everyday use in any language, although the Guinness Book of World Records (1998) recognizes the somewhat longer "saippuakivikauppias" (dealer in lye) as the longest palindromic word. On the other hand, the Dutch palindrome "Edelstaalplaatslede" (steel tray in an oven) is just as long, "Koortsmeetsysteemstrook" is longer, and yet even longer examples exist in Finnish itself. So the title of longest single-word palindrome would still seem to be up for grabs. Here now is a palindrome sampler from twelve different languages:

Nisumaa oli isasi ilo aamusin.
Isa, ala myy myymalaasi.
Reit nie tot ein Tier.
Erika feuert nur untreue Fakire.
νιψον ανομηματα μη μοναν οψιν
En af dem der tit red med fane.
"Mooie zeden in Ede", zei Oom.
I topi non avevano nipoti.
e' li' Bari, mirabile.
Ai lati d'Italia.
Llad dafad dall.
A mala nada na lama.
Elu par cette crapule.
Engage le jeu que je le gagne.
Dábale arroz a la zorra el abad.
Roma tibi subito motibus ibit amor.
In girum imus nocte, et consumimur igni.
Ni talar bra latin.

Other types of palindromes are also of interest. Contrasting with the traditional "letter-unit" palindromes above are the so-called "word-unit" palindromes such as

All for one and one for all.
Fair is foul, and foul is fair.
So patient a doctor to doctor a patient so.
Stout and bitter porter drinks porter, bitter and stout.
Girl, bathing on Bikini, eyeing boy,
finds
boy eyeing bikini on bathing girl.
Bob: did Anna peep? Anna: did Bob?

and phonetic palindromes such as "Ominous cinema". Some have even gone so far as to construct palindromic poems and even palindromic novels, in both letter-unit and word-unit forms, although these tend to lose cogency as the length increases. Notable among these efforts is the work of Georges Perec [7] and the literary group Oulipo (Ouvroir de Littérature Potentielle) [67].

Perhaps the most dramatic development in the field of palindromology is the recent discovery [71], [77], [87] that the human male Y chromosome contains an approximately 3-million letter palindrome in its DNA sequence, making this probably the longest known naturally-occurring palindrome.

Visually symmetric realizations of words and phrases can be especially appealing to those who delight in palindromes. For a cornucopia of calligraphic analogs of palindromes and T-palindromes see the books of Scott Kim [44] and John Langdon [52].

Chapter 7

Structured Linearizations

Having now established the basic properties of palindromic and alternating polynomials, and indicated their applicability to a variety of practical problems, the main goal of this chapter is to show how to find *structured linearizations* for these classes of structured polynomial. Our strategy is to search for these structured linearizations in the vector spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$, pencil spaces that were originally designed to provide large arenas fertile enough to contain linearizations that reflect additional structure in P, but small enough that pencils $L(\lambda) \in \mathbb{L}_j(P)$ still share salient features of the companion forms $C_j(\lambda)$.

In this endeavor we will restrict ourselves to the $\star$-variants of these structures. There are two reasons for this, one mathematical and one practical. The practical reason is that "pure" palindromic, anti-palindromic, even, and odd matrix polynomials just don't seem to show up in any applications. The mathematical reason is more cogent; in general, these "pure" structures cannot be linearized in a structure preserving way.

Consider, for example, a regular $n \times n$ palindromic polynomial $P(\lambda)$ of degree $k \geq 2$. By [31, Theorem 1.7] a pencil $L(\lambda)$ can only be a linearization for an $n \times n$ matrix polynomial if the geometric multiplicity of each eigenvalue of $L(\lambda)$ is less than or equal to n. On the other hand, any palindromic linearization must be of the form $L(\lambda) = \lambda Z + Z$ for some matrix Z; for such a pencil the eigenvalue -1 has geometric multiplicity kn, thus ruling out any palindromic linearization. What about T-palindromic linearizations for $P(\lambda)$? In the quadratic case $P(\lambda) = \lambda^2 A + \lambda B + A$, and a calculation similar to the one in Example 7.1.1 shows that A and B must both be symmetric or both skew-symmetric (so that P is T-(anti)-palindromic as well as palindromic) in order for any $L(\lambda) \in \mathbb{L}_1(P)$ of the form $\lambda Z + Z^T$ to exist. Thus general palindromic matrix polynomials do not admit T-palindromic linearizations in $\mathbb{L}_1(P)$. Analogous arguments exclude structured linearizations for general anti-palindromic, even, and odd polynomials.

Recall from Theorem 2.2.7 that simply being an element of $\mathbb{L}_1(P)$ is already almost enough to guarantee being a linearization of P. Hence we begin our search for structured linearizations by first trying to demonstrate the existence of structured pencils in $\mathbb{L}_1(P)$. In later sections we will see what (if any) extra conditions are needed to guarantee that these structured pencils are indeed linearizations for P.

7.1 A Simple Example

In Chapter 5 we found that structured pencils in $\mathbb{L}_1(P)$ for symmetric and Hermitian polynomials P were all contained in the space $\mathbb{DL}(P) := \mathbb{L}_1(P) \cap \mathbb{L}_2(P)$. Given that pencils from $\mathbb{DL}(P)$ are always block-symmetric [Thm. 3.4.2], this is perhaps not so surprising. However, this same block-symmetry of $\mathbb{DL}(P)$-pencils makes it too much to expect that pencils reflecting $\star$-palindromic or $\star$-alternating structure can also be found in $\mathbb{DL}(P)$. So instead we regard all of $\mathbb{L}_1(P)$ as the search space, and begin our investigation with a simple example of low degree, worked out from first principles.

Example 7.1.1. Consider the T-palindromic matrix polynomial $P(\lambda) = \lambda^2 A + \lambda B + A^T$, where $A, B \in \mathbb{F}^{n\times n}$, $B^T = B$ and $A \neq 0$. Our goal is to construct a pencil $L(\lambda)$ in $\mathbb{L}_1(P)$ with the same (T-palindromic) structure as P; to avoid trivialities we also insist that $L(\lambda)$ should have a nonzero right ansatz vector $v = [v_1, v_2]^T \in \mathbb{F}^2$. This means that $L(\lambda)$ must be of the form

$$L(\lambda) = \lambda Z + Z^T =: \lambda \begin{bmatrix} D & E \\ F & G \end{bmatrix} + \begin{bmatrix} D^T & F^T \\ E^T & G^T \end{bmatrix} \quad \text{for some } \ D, E, F, G \in \mathbb{F}^{n\times n}.$$

Since $L(\lambda) \in \mathbb{L}_1(P)$, the equivalence in Lemma 2.1.4 implies that we can rewrite this using the column shifted sum $\boxplus\!\!\rightarrow$ as

$$Z \boxplus\!\!\rightarrow Z^T = \begin{bmatrix} D & E + D^T & F^T \\ F & G + E^T & G^T \end{bmatrix} = \begin{bmatrix} v_1 A & v_1 B & v_1 A^T \\ v_2 A & v_2 B & v_2 A^T \end{bmatrix}.$$

Equating corresponding blocks in the first and last columns, we obtain $D = v_1 A$, $F = v_2 A = v_1 A$, and $G = v_2 A$. This forces $v_1 = v_2$, since $A \neq 0$ by assumption. From either block of the middle column we see that $E = v_1(B - A^T)$; with this choice for E all the equations are consistent, thus yielding

$$L(\lambda) = \lambda Z + Z^T = v_1 \left(\lambda \begin{bmatrix} A & B - A^T \\ A & A \end{bmatrix} + \begin{bmatrix} A^T & A^T \\ B - A & A^T \end{bmatrix} \right). \tag{7.1.1}$$

This gives us a T-palindromic pencil in $\mathbb{L}_1(P)$ with right ansatz vector $v = v_1[\,1, 1\,]^T$.

Example 7.1.1 illustrates three important properties that turn out to hold more generally. First, the choice of right ansatz vector v for which the corresponding $L(\lambda) \in \mathbb{L}_1(P)$ is T-palindromic is restricted to ones that are themselves palindromic. On the other hand, once a palindromic right ansatz vector v is chosen, the pencil $L(\lambda) \in \mathbb{L}_1(P)$ is uniquely determined by insisting that it be T-palindromic. Finally, although $L(\lambda)$ is itself not in $\mathbb{DL}(P)$, it is easily converted into a $\mathbb{DL}(P)$-pencil simply by interchanging the first and second block rows of $L(\lambda)$, or equivalently, by premultiplying by $R_2 \otimes I$ where R_2 is the 2×2 reverse identity as in (6.1.4). This yields the pencil

$$(R_2 \otimes I)L(\lambda) = v_1 \left(\lambda \begin{bmatrix} A & A \\ A & B - A^T \end{bmatrix} + \begin{bmatrix} B - A & A^T \\ A^T & A^T \end{bmatrix} \right),$$

which is readily confirmed to be in $\mathbb{DL}(P)$ with ansatz vector $v = v_1[\,1, 1\,]^T$ by using column and row shifted sums.

We now state a theorem that generalizes these observations about Example 7.1.1. Here $R = R_k$ as in (6.1.4) with $k = \deg P$.

Theorem 7.1.2. *Let $P(\lambda)$ be a T-palindromic matrix polynomial and $L(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector v. Then the pencil $L(\lambda)$ is T-palindromic if and only if $Rv = v$ and $(R \otimes I)L(\lambda) \in \mathbb{DL}(P)$ with ansatz vector Rv. Moreover, for* any $v \in \mathbb{F}^k$ *satisfying $Rv = v$ there exists a unique pencil $L(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector v and T-palindromic structure.*

The proof of this theorem is deferred to the next section, where it is subsumed under an even more general result [Thm. 7.2.1] that encompasses all eight combinations of $\star$-(anti)-palindromic polynomial P with $\star$-(anti)-palindromic pencil in $\mathbb{L}_1(P)$.

7.2 Existence of Structured Pencils in $\mathbb{L}_1(P)$

For a $\star$-palindromic or $\star$-alternating polynomial P it is natural to try to find a linearization with the *same* structure as P. In this section we begin that search by characterizing the pencils in $\mathbb{L}_1(P)$ with the same structure as P. From the point of view of numerical analysis, however, one of the most important reasons for using a structure-preserving method is to preserve spectral symmetries. But we see in Table 6.2.1 that for each structure under consideration there is also an "anti" version of that structure with *the same spectral symmetry*. Thus it makes sense to try to linearize a structured polynomial with an "anti-structured" pencil as well as with a structured one; so in this section we also characterize the pencils in $\mathbb{L}_1(P)$ having the anti-structure of P.

We begin with the $\star$-palindromic structures, showing that, just as in Theorem 7.1.2, there is only a restricted class of admissible right ansatz vectors v that can support a structured or anti-structured pencil in $\mathbb{L}_1(P)$. In each case the restrictions on the vector v can be concisely described using the reverse identity $R = R_k$ as defined in (6.1.4). For the $\star$-alternating structures there are analogous results where the restrictions on the admissible right ansatz vectors v are described using the diagonal alternating-signs matrix $\Sigma = \Sigma_k$ as in (6.1.4).

Theorem 7.2.1 (Existence/Uniqueness of Structured Pencils, Part 1). *Suppose the matrix polynomial $P(\lambda)$ is $\star$-palindromic or $\star$-anti-palindromic. Then for pencils $L(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector v, conditions (i) and (ii) in Table* 7.2.1 *are equivalent. Moreover, for* any $v \in \mathbb{F}^k$ *satisfying one of the admissibility conditions for v in (ii), there exists a unique pencil $L(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector v and the corresponding structure in (i).*

Proof. We consider all eight cases simultaneously. Let $P(\lambda)$ be $\star$-palindromic or $\star$-anti-palindromic, so that $\operatorname{rev} P^\star(\lambda) = \chi_P P(\lambda)$ for $\chi_P = \pm 1$.

"(*i*) $\Rightarrow$ (*ii*)": Condition (i) means that $L(\lambda)$ satisfies $\operatorname{rev} L^\star(\lambda) = \chi_L L(\lambda)$ for $\chi_L = \pm 1$, and since $L(\lambda) \in \mathbb{L}_1(P)$,

$$L(\lambda)(\Lambda \otimes I) = v \otimes P(\lambda). \tag{7.2.1}$$

Taking the reversal of both sides of (7.2.1), and noting that $R\Lambda = \operatorname{rev}\Lambda$, we have

$$\operatorname{rev} L(\lambda)(R \otimes I)(\Lambda \otimes I) = \operatorname{rev} L(\lambda)\big((\operatorname{rev}\Lambda) \otimes I\big) = v \otimes \operatorname{rev} P(\lambda)\,.$$

Table 7.2.1: Admissible ansatz vectors for structured pencils (palindromic structures)

Structure of $P(\lambda)$	Equivalent conditions	
	(i) $L(\lambda)$ is	(ii) $(R \otimes I)L(\lambda) \in \mathbb{DL}(P)$ with ansatz vector Rv and
T-palindromic	T-palindromic	$Rv = v$
	T-anti-palindromic	$Rv = -v$
T-anti-palindromic	T-palindromic	$Rv = -v$
	T-anti-palindromic	$Rv = v$
$*$-palindromic	$*$-palindromic	$Rv = \overline{v}$
	$*$-anti-palindromic	$Rv = -\overline{v}$
$*$-anti-palindromic	$*$-palindromic	$Rv = -\overline{v}$
	$*$-anti-palindromic	$Rv = \overline{v}$

Now applying the adjoint $\star$ to both sides, we obtain

$$(\varLambda^\star \otimes I)(R \otimes I)\,\mathrm{rev}\,L^\star(\lambda^\star) = v^\star \otimes \mathrm{rev}\,P^\star(\lambda^\star)\,,$$

or equivalently,

$$(\varLambda^\star) \otimes I)(R \otimes I)L(\lambda^\star) = (\chi_P \chi_L v^\star) \otimes P(\lambda^\star)\,, \tag{7.2.2}$$

since $L(\lambda)$ and $P(\lambda)$ are either $\star$-palindromic or $\star$-anti-palindromic. Then using the fact that (7.2.2) is an identity, we replace $\lambda^\star$ by λ to obtain

$$(\varLambda^T \otimes I)(R \otimes I)L(\lambda) = (\chi_P \chi_L v^\star) \otimes P(\lambda)\,, \tag{7.2.3}$$

thus showing the pencil $(R \otimes I)L(\lambda)$ to be an element of $\mathbb{L}_2(P)$ with left ansatz vector $w = \chi_P \chi_L (v^\star)^T$. On the other hand, multiplying (7.2.1) on the left by $R \otimes I$ yields

$$(R \otimes I)L(\lambda)(\varLambda \otimes I) = (Rv) \otimes P(\lambda)\,, \tag{7.2.4}$$

so $(R \otimes I)L(\lambda)$ is also in $\mathbb{L}_1(P)$ with right ansatz vector Rv. Thus $(R \otimes I)L(\lambda)$ is in $\mathbb{DL}(P) = \mathbb{L}_1(P) \cap \mathbb{L}_2(P)$, and from Theorem 3.4.2 the equality of right and left ansatz vectors implies that

$$Rv = \chi_P \chi_L (v^\star)^T = \begin{cases} \chi_P \chi_L v & \text{when } \star = T\,, \\ \chi_P \chi_L \overline{v} & \text{when } \star = *\,. \end{cases}$$

All eight variants of condition (ii) now follow.

"$(ii) \Rightarrow (i)$": Since $(R \otimes I)L(\lambda)$ is in $\mathbb{DL}(P)$ with ansatz vector Rv, we have

$$\begin{aligned} (R \otimes I)L(\lambda)(\varLambda \otimes I) &= (Rv) \otimes P(\lambda)\,, & (7.2.5)\\ \big((\varLambda^T R) \otimes I\big)L(\lambda) = (\varLambda^T \otimes I)(R \otimes I)L(\lambda) &= (Rv)^T \otimes P(\lambda)\,. & (7.2.6) \end{aligned}$$

Applying the adjoint $\star$ to both ends of (7.2.6) gives

$$L^\star(\lambda^\star)\big((R(\Lambda^T)^\star)\otimes I\big) = R(v^T)^\star \otimes P^\star(\lambda^\star)\,,$$

or equivalently

$$L^\star(\lambda)\big((R\Lambda)\otimes I\big) = R(v^T)^\star \otimes P^\star(\lambda)\,. \tag{7.2.7}$$

Note that all cases of condition (ii) may be expressed in the form $R(v^T)^\star = \varepsilon\chi_{_P} v$, where $\varepsilon = \pm 1$. Then taking the reversal of both sides in (7.2.7) and using $R\Lambda = \mathrm{rev}\,\Lambda$ we obtain

$$\mathrm{rev}\,L^\star(\lambda)(\Lambda\otimes I) = (\varepsilon\chi_{_P} v)\otimes \mathrm{rev}\,P^\star(\lambda) = (\varepsilon v)\otimes P(\lambda)\,,$$

and after multiplying by $\varepsilon(R\otimes I)$,

$$\varepsilon(R\otimes I)\mathrm{rev}\,L^\star(\lambda)(\Lambda\otimes I) = (Rv)\otimes P(\lambda). \tag{7.2.8}$$

Thus we see that the pencil $\varepsilon(R\otimes I)\mathrm{rev}\,L^\star(\lambda)$ is in $\mathbb{L}_1(P)$ with right ansatz vector Rv. Now starting over again from identity (7.2.5) and taking the adjoint $\star$ of both sides, we obtain by analogous reasoning that

$$\begin{aligned}
&(R\otimes I)L(\lambda)(\Lambda\otimes I) = (Rv)\otimes P(\lambda)\\
&\iff (\Lambda^T\otimes I)L^\star(\lambda)(R\otimes I) = (v^\star R)\otimes P^\star(\lambda) = \big(v^\star\otimes P^\star(\lambda)\big)(R\otimes I)\\
&\iff (\Lambda^T\otimes I)L^\star(\lambda) = v^\star\otimes P^\star(\lambda)\\
&\iff (\mathrm{rev}\,\Lambda^T\otimes I)\,\mathrm{rev}\,L^\star(\lambda) = v^\star\otimes \mathrm{rev}\,P^\star(\lambda)\\
&\iff (\Lambda^T R\otimes I)\,\mathrm{rev}\,L^\star(\lambda) = (\varepsilon\chi_{_P} Rv)^T\otimes \mathrm{rev}\,P^\star(\lambda) = (\varepsilon Rv)^T\otimes P(\lambda)\\
&\iff (\Lambda^T\otimes I)\Big(\varepsilon(R\otimes I)\mathrm{rev}\,L^\star(\lambda)\Big) = (Rv)^T\otimes P(\lambda).
\end{aligned} \tag{7.2.9}$$

Thus the pencil $\varepsilon(R\otimes I)\mathrm{rev}\,L^\star(\lambda)$ is also in $\mathbb{L}_2(P)$ with left ansatz vector Rv. Taken together, (7.2.8) and (7.2.9) show that $\varepsilon(R\otimes I)\,\mathrm{rev}\,L^\star(\lambda)$ is in $\mathbb{DL}(P)$ with ansatz vector Rv. But $(R\otimes I)L(\lambda)$ is also in $\mathbb{DL}(P)$ with exactly the same ansatz vector, so the uniqueness property of Theorem 3.4.2 for $\mathbb{DL}(P)$-pencils implies that

$$\varepsilon(R\otimes I)\,\mathrm{rev}\,L^\star(\lambda) \,\equiv\, (R\otimes I)L(\lambda)\,,$$

or equivalently $\varepsilon\,\mathrm{rev}\,L^\star(\lambda) = L(\lambda)$. Hence $L(\lambda)$ is $\star$-palindromic or $\star$-anti-palindromic, depending on the parameter ε, which implies all the variants of condition (i) in Table 7.2.1.

Finally, the existence and uniqueness of a structured pencil $L(\lambda)$ corresponding to any admissible right ansatz vector v follows directly from the existence and uniqueness in Theorem 3.4.2 of the $\mathbb{DL}(P)$-pencil $(R\otimes I)L(\lambda)$ for the ansatz vector Rv. ∎

We next present the analog of Theorem 7.2.1 for $\star$-even and $\star$-odd polynomials. Here the matrix $\Sigma = \Sigma_k$ is the diagonal matrix of alternating signs as defined in (6.1.4).

Theorem 7.2.2 (Existence/Uniqueness of Structured Pencils, Part 2). *Suppose the matrix polynomial $P(\lambda)$ is $\star$-even or $\star$-odd. Then for pencils $L(\lambda)\in\mathbb{L}_1(P)$ with right ansatz vector v, conditions (i) and (ii) in Table 7.2.2 are equivalent. Moreover, for any $v\in\mathbb{F}^k$ satisfying one of the admissibility conditions for v in (ii), there exists a unique pencil $L(\lambda)\in\mathbb{L}_1(P)$ with right ansatz vector v and the corresponding structure in (i).*

Table 7.2.2: Admissible ansatz vectors for structured pencils (alternating structures)

Structure of $P(\lambda)$	Equivalent conditions	
	(i) $L(\lambda)$ is	(ii) $(\Sigma \otimes I)L(\lambda) \in \mathbb{DL}(P)$ with ansatz vector Σv and
T-even	T-even	$\Sigma v = v$
	T-odd	$\Sigma v = -v$
T-odd	T-even	$\Sigma v = -v$
	T-odd	$\Sigma v = v$
$*$-even	$*$-even	$\Sigma v = \overline{v}$
	$*$-odd	$\Sigma v = -\overline{v}$
$*$-odd	$*$-even	$\Sigma v = -\overline{v}$
	$*$-odd	$\Sigma v = \overline{v}$

Proof. The proof proceeds in a completely analogous fashion to the proof of Theorem 7.2.1, with only two differences. The first is that in steps of the proof where we took the reversal of two sides of an equation in Theorem 7.2.1, instead we now simply replace λ by $-\lambda$. Observe that replacing λ by $-\lambda$ in Λ has the same effect as premultiplying it by Σ, that is $\Sigma\Lambda = \Lambda(-\lambda) = [(-\lambda)^{k-1}, \ldots, -\lambda, 1]^T$. The other difference is that multiplications by $R \otimes I$ are replaced with multiplications by $\Sigma \otimes I$. ∎

The connections established in Theorems 7.2.1 and 7.2.2 between $\star$-structured pencils and $\mathbb{DL}(P)$-pencils have two important consequences. The first is an efficient procedure, to be discussed in section 7.3.1, for explicitly constructing these $\star$-structured pencils from the $\mathbb{DL}(P)$ standard basis developed in section 3.3.2. The second is a way to tell which of these structured pencils are actually *linearizations* for P (see section 7.4 for details).

7.3 Construction of Structured Pencils

Having established the conditions under which $\star$-structured pencils exist, we now describe two methods for constructing them. The first method is based on the connection between these structured pencils and $\mathbb{DL}(P)$-pencils, while the second goes back to first principles to build structured pencils via the systematic interweaving of shifted sums and the invocation of $\star$-structure.

7.3.1 Construction Using the Standard Basis for $\mathbb{DL}(P)$

As we have seen in Theorems 7.2.1 and 7.2.2, $\star$-structured pencils in $\mathbb{L}_1(P)$ are strongly related to elements of the space $\mathbb{DL}(P)$. In particular, if $L(\lambda) \in \mathbb{L}_1(P)$ is $\star$-structured with right ansatz vector v, then either $\widetilde{L}(\lambda) = (R \otimes I)L(\lambda)$ or $\widetilde{L}(\lambda) = (\Sigma \otimes I)L(\lambda)$ is in $\mathbb{DL}(P)$ with ansatz vector $w = Rv$ or $w = \Sigma v$, respectively, depending on whether palindromic or alternating structure is present. This observation leads to the following procedure for the construction of structured pencils:

(1) Choose a right ansatz vector $v \in \mathbb{F}^k$ that is admissible for the desired type of $\star$-structure, i.e., one satisfying the appropriate condition in Table 7.2.1 or 7.2.2.

(2) Let w be Rv or Σv, respectively, and construct the unique pencil $\widetilde{L}(\lambda) \in \mathbb{DL}(P)$ with ansatz vector w.

(3) Premultiply $\widetilde{L}(\lambda)$ with $R \otimes I$ or $\Sigma \otimes I$, respectively, to recover the desired $\star$-structured pencil $L(\lambda) \in \mathbb{L}_1(P)$ with right ansatz vector v. (Recall that $R^2 = \Sigma^2 = I$.)

All that remains is to observe that step (2) may be carried out concretely and explicitly using the standard basis for $\mathbb{DL}(P)$ described in section 3.3.2. Recall from Theorem 3.3.2 that the pencil $\lambda X_m - X_{m-1}$, where X_m is defined in (3.3.6), (3.3.7), and (3.3.8), is the mth standard basis pencil for $\mathbb{DL}(P)$ corresponding to the ansatz vector e_m. Then the $\mathbb{DL}(P)$-pencil with ansatz vector w is just the linear combination

$$\widetilde{L}(\lambda) \;=\; \sum_{j=1}^{k} w_j\bigl(\lambda X_j - X_{j-1}\bigr) \;=\; \lambda\sum_{j=1}^{k} w_j X_j - \sum_{j=1}^{k} w_j X_{j-1}\,. \tag{7.3.1}$$

It should be emphasized here that any structured pencil $L(\lambda) \in \mathbb{L}_1(P)$ obtained by this procedure is only a *potential* linearization, because only regular pencils in $\mathbb{L}_1(P)$ are linearizations for $P(\lambda)$. (See the Strong Linearization Theorem, Thm. 2.2.3) We will return to the issue of determining which of these structured pencils are actually linearizations in section 7.4.

7.3.2 Structured Pencils via Shifted Sums

We now discuss an alternative approach, based directly on the use of shifted sums, for the construction of $\star$-structured pencils in $\mathbb{L}_1(P)$. Indeed, it is worth noting that this was the *original* method for building these structured pencils. We begin with an illustration of this approach using an example of low degree.

Suppose we start with a T-palindromic polynomial $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda B^T + A^T$, and our aim is to construct a T-palindromic pencil $L(\lambda) \in \mathbb{L}_1(P)$. By Theorem 7.2.1 the corresponding right ansatz vector v must satisfy $Rv = v$, so let us choose $v = [1, -1, 1]^T$. Then with $L(\lambda) := \lambda Z + Z^T$, the shifted sum $Z \boxplus\!\!\rightarrow Z^T$ must by Lemma 2.1.4 be

$$Z \boxplus\!\!\rightarrow Z^T = v \otimes [\, A \quad B \quad B^T \quad A^T\,] = \begin{bmatrix} A & B & B^T & A^T \\ -A & -B & -B^T & -A^T \\ A & B & B^T & A^T \end{bmatrix}. \tag{7.3.2}$$

By the definition of the shifted sum, the first block column of Z and the last block column of Z^T are now uniquely determined. Hence

$$\lambda Z + Z^T = \lambda \begin{bmatrix} A & * & * \\ -A & * & * \\ A & * & * \end{bmatrix} + \begin{bmatrix} * & * & A^T \\ * & * & -A^T \\ * & * & A^T \end{bmatrix}, \tag{7.3.3}$$

where $*$ represents $n \times n$ blocks yet to be determined. We now continue by alternately using the fact that $L(\lambda)$ is T-palindromic and that $L(\lambda)$ is in $\mathbb{L}_1(P)$. Thus, observing that the second matrix in (7.3.3) is just the transpose of the first one, we obtain

$$\lambda Z + Z^T = \lambda \begin{bmatrix} A & * & * \\ -A & * & * \\ A & -A & A \end{bmatrix} + \begin{bmatrix} A^T & -A^T & A^T \\ * & * & -A^T \\ * & * & A^T \end{bmatrix}.$$

Then invoke (7.3.2) again, which forces

$$\lambda Z + Z^T = \lambda \begin{bmatrix} A & B - A^T & B^T + A^T \\ -A & * & * \\ A & -A & A \end{bmatrix} + \begin{bmatrix} A^T & -A^T & A^T \\ * & * & -A^T \\ B + A & B^T - A & A^T \end{bmatrix}.$$

The two matrices of the pencil are still transposes of one another, so this now implies

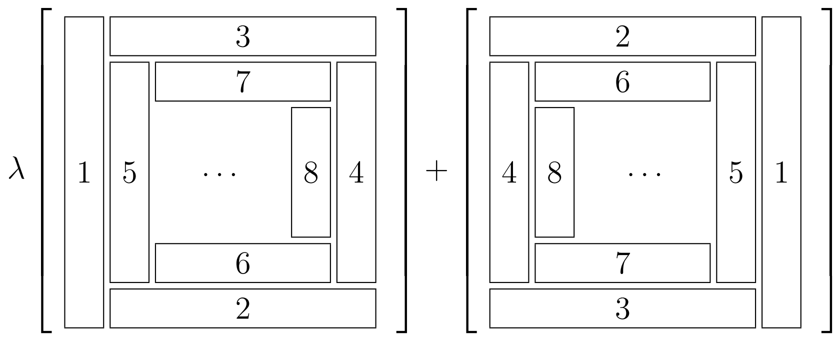
$$\lambda Z + Z^T = \lambda \begin{bmatrix} A & B - A^T & B^T + A^T \\ -A & * & B - A^T \\ A & -A & A \end{bmatrix} + \begin{bmatrix} A^T & -A^T & A^T \\ B^T - A & * & -A^T \\ B + A & B^T - A & A^T \end{bmatrix}.$$

Using (7.3.2) once more, we finally obtain

$$\lambda Z + Z^T = \lambda \begin{bmatrix} A & B - A^T & B^T + A^T \\ -A & A - B^T - B & B - A^T \\ A & -A & A \end{bmatrix} + \begin{bmatrix} A^T & -A^T & A^T \\ B^T - A & A^T - B - B^T & -A^T \\ B + A & B^T - A & A^T \end{bmatrix},$$

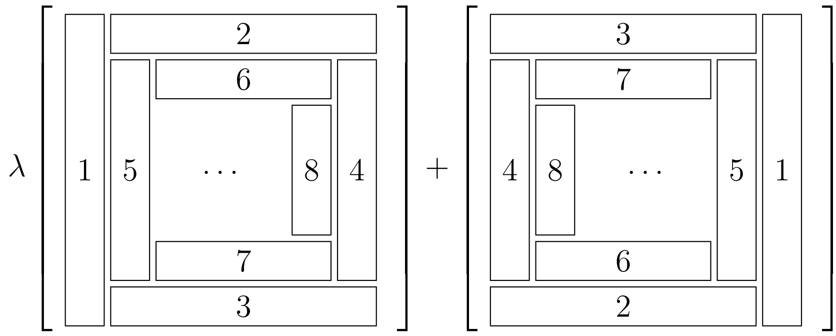
thus completing the construction.

More generally, suppose $P(\lambda) = \sum_{j=1}^{k} \lambda^j A_j$ is a $\star$-palindromic or $\star$-anti-palindromic matrix polynomial and we want $L(\lambda) \in \mathbb{L}_1(P)$ to have the form $\lambda Z + \varepsilon Z^\star$ with $\varepsilon = \pm 1$. Then we can construct such a pencil in block-column/block-row-wise fashion following the order displayed in (7.3.4). Here, each *panel* (that is, each portion of a block column or block row) labelled with an odd number is determined by using information from the shifted sum property $Z \boxplus\!\!\rightarrow (\varepsilon Z^\star) = v \otimes \begin{bmatrix} A_k & \cdots & A_0 \end{bmatrix}$, and each panel labelled with an even number is determined by requiring that $L(\lambda)$ is, depending on ε, either $\star$-palindromic or $\star$-anti-palindromic, respectively.



(7.3.4)

The construction of $\star$-even or $\star$-odd pencils for $\star$-even or $\star$-odd matrix polynomials is carried out along similar lines, following the pattern in (7.3.5).



(7.3.5)

Again, each panel labelled with an odd number is constructed using information from the fact that the desired pencil is to be in $\mathbb{L}_1(P)$, while each panel labelled with an even number is constructed to maintain the $\star$-even or $\star$-odd structure of the pencil.

Because of the existence/uniqueness theorems for $\star$-structured pencils proved in section 7.2, it follows that these procedures will work in general, provided that an admissible right ansatz vector v for the given combination of structure in P and L, i.e. one that satisfies the appropriate restriction described in Table 7.2.1 or 7.2.2, is chosen to start the process off with.

A complete list of structured pencils for structured polynomials of degree two is given in Table 7.7.1. Note that we do not include either $*$-odd or $*$-anti-palindromic matrix polynomials in this list, because multiplication by $i \in \mathbb{C}$ immediately transforms them into $*$-even and $*$-palindromic matrix polynomials, respectively. Some selected structured pencils for $\star$-palindromic and $\star$-even matrix polynomials of degree three are given in Tables 7.7.2 and 7.7.3.

7.4 Which Structured Pencils are Linearizations?

Now that we know when structured pencils in $\mathbb{L}_1(P)$ exist and how to construct them, the one fundamental issue that remains is to determine which ones (if any) are actually linearizations for P. Because of the intimate connections established in section 7.2 between these structured pencils and $\mathbb{DL}(P)$-pencils, the linearization question is easy to settle by using the eigenvalue exclusion theorem from Chapter 4. We recall that theorem again for the convenience of the reader.

Theorem 4.3.1 (Eigenvalue Exclusion Theorem).
Suppose that $P(\lambda)$ is a regular matrix polynomial and $L(\lambda)$ is in $\mathbb{DL}(P)$ with nonzero ansatz vector v. Then $L(\lambda)$ is a (strong) linearization for $P(\lambda)$ if and only if no root of the v*-polynomial $\mathsf{p}(x\,;v)$ is an eigenvalue of $P(\lambda)$. (Note that this statement includes ∞ as one of the possible roots of $\mathsf{p}(x\,;v)$ or possible eigenvalues of $P(\lambda)$.)*

From this result we can now quickly deduce the following theorem.

Theorem 7.4.1 (Structured Linearization Theorem).
Suppose the regular matrix polynomial $P(\lambda)$ and the nonzero pencil $L(\lambda) \in \mathbb{L}_1(P)$ have one of the sixteen combinations of $\star$-structure considered in Tables 7.2.1 *and*

7.2.2. *Let v be the nonzero right ansatz vector of $L(\lambda)$, and let*

$$w = \begin{cases} Rv & \text{if } P \text{ is } \star\text{-palindromic or } \star\text{-anti-palindromic}, \\ \Sigma v & \text{if } P \text{ is } \star\text{-even or } \star\text{-odd}. \end{cases}$$

Then $L(\lambda)$ is a (strong) linearization for $P(\lambda)$ if and only if no root of the v-polynomial $\mathsf{p}(x\,;w)$ is an eigenvalue of $P(\lambda)$.

Proof. For all eight $\star$-palindromic structure combinations, it was shown in Theorem 7.2.1 that $(R\otimes I)L(\lambda)$ is in $\mathbb{DL}(P)$ with ansatz vector Rv. Similarly for the eight $\star$-alternating structure combinations it was shown in Theorem 7.2.2 that $(\Sigma\otimes I)L(\lambda)$ is in $\mathbb{DL}(P)$ with ansatz vector Σv. Since clearly $L(\lambda)$ is a linearization for $P(\lambda)$ if and only if $(R\otimes I)L(\lambda)$ or $(\Sigma\otimes I)L(\lambda)$ is, the desired result follows immediately from the eigenvalue exclusion theorem. ∎

Note that the tables of structured pencils at the end of this chapter also include the corresponding linearization conditions obtained from Theorem 7.4.1. We illustrate with an example from Table 7.7.1.

Example 7.4.2. Suppose the T-palindromic polynomial $P(\lambda) = \lambda^2 A + \lambda B + A^T$ from Example 7.1.1 is regular. Theorem 7.1.2 restricts the admissible right ansatz vectors $v \in \mathbb{F}^2$ of a T-palindromic pencil $L(\lambda) \in \mathbb{L}_1(P)$ to those that satisfy $Rv = v$, or equivalently, $v = (v_1, v_1)^T$. We see from Theorem 7.4.1 that such an $L(\lambda)$ will be a strong linearization for $P(\lambda)$ if and only if none of the roots of the v-polynomial $\mathsf{p}(x\,;Rv) = v_1x + v_1$ are eigenvalues of $P(\lambda)$, that is, if and only if -1 is not an eigenvalue of $P(\lambda)$. On the other hand, a T-anti-palindromic pencil $\widetilde{L}(\lambda) \in \mathbb{L}_1(P)$ will be a linearization for P if and only if $\lambda = 1$ is not an eigenvalue of $P(\lambda)$. This is because every admissible right ansatz vector for $\widetilde{L}(\lambda)$ is constrained by Theorem 7.2.1 to be of the form $\widetilde{v} = [\,v_1, -v_1\,]^T$, forcing $\mathsf{p}(x\,;R\widetilde{v}) = -v_1x + v_1$, with only $+1$ as a root.

Although Theorem 7.4.1 settles the issue about which structured pencils are linearizations, there is one further aspect of structure in this story. Theorems 7.2.1 and 7.2.2 put restrictions on the admissible right ansatz vectors of structured pencils, which means that these vectors are themselves structured, and consequently so are the scalar v-polynomials associated with them. What is interesting is that the structure of the right ansatz vector and v-polynomial *parallels* the structure of the matrix polynomial $P(\lambda)$ and linearizing pencil $L(\lambda) \in \mathbb{L}_1(P)$.

Suppose, as in Example 7.4.2, that $P(\lambda)$ is T-palindromic and we want a T-palindromic linearization $L(\lambda) \in \mathbb{L}_1(P)$. Then by Theorem 7.2.1 any admissible right ansatz vector satisfies $Rv = v$, and so has components that read the same forwards or backwards. Thus v is itself palindromic, and the corresponding v-polynomial $\mathsf{p}(x\,;Rv)$ is T-palindromic (and also palindromic, since it is a scalar polynomial).

Theorems 7.2.1 and 7.2.2 imply that analogous parallels in structure hold for other combinations of $\star$-structures in P and L and the relevant v-polynomial $\mathsf{p}(x\,;Rv)$ or $\mathsf{p}(x\,;\Sigma v)$; for convenience these are listed together in Table 7.4.1.

Table 7.4.1: Parallelism of Structures

$P(\lambda)$	$L(\lambda) \in \mathbb{L}_1(P)$	v-polynomial	$P(\lambda)$	$L(\lambda) \in \mathbb{L}_1(P)$	v-poly.
$\star$-palin.	$\star$-palin.	$\star$-palin.	$\star$-even	$\star$-even	$\star$-even
	$\star$-anti-palin.	$\star$-anti-palin.		$\star$-odd	$\star$-odd
$\star$-anti-palin.	$\star$-palin.	$\star$-anti-palin.	$\star$-odd	$\star$-even	$\star$-odd
	$\star$-anti-palin.	$\star$-palin.		$\star$-odd	$\star$-even

7.5 When Pairings Degenerate

The parallel of structures between matrix polynomial, $\mathbb{L}_1(P)$-pencil, and v-polynomial (see Table 7.4.1) is aesthetically very pleasing: structure in a v-polynomial forces a pairing of its roots as in Theorem 6.2.1 which is always of the *same qualitative type* as the eigenvalue pairing present in the original structured matrix polynomial. However, it turns out that this root pairing can sometimes be an obstruction to the existence of any structured linearization in $\mathbb{L}_1(P)$ at all.

Using an argument based mainly on the very simple form of admissible right ansatz vectors when $k = 2$, we saw in Example 7.4.2 that a quadratic T-palindromic matrix polynomial having *both* $+1$ and -1 as eigenvalues cannot have a structured linearization in $\mathbb{L}_1(P)$: the presence of -1 in the spectrum precludes the existence of a T-palindromic linearization, while the eigenvalue $+1$ excludes T-anti-palindromic linearizations. We now show that this difficulty is actually a consequence of root pairing, and therefore can also occur for higher degree polynomials.

Whenever $P(\lambda)$ has even degree, all right ansatz vectors of $\mathbb{L}_1(P)$-pencils have even length, and hence the corresponding v-polynomials all have an odd number of roots (counting multiplicities and including ∞). Root pairing then forces at least one root of every v-polynomial to lie in a subset of $\mathbb{C}$ where this pairing "degenerates". For $*$-palindromic and $*$-alternating polynomials the pairings are $(\lambda, 1/\overline{\lambda}\,)$ and $(\lambda, -\overline{\lambda}\,)$, so the degeneration sets are the unit circle and the imaginary axis (including ∞), respectively. By contrast the pairings for T-palindromic and T-alternating polynomials are $(\lambda, 1/\lambda)$ and $(\lambda, -\lambda)$, so in these cases the degeneration sets are the finite sets $\{-1, +1\}$ and $\{0, \infty\}$, respectively. Thus for any T-(anti)-palindromic matrix polynomial $P(\lambda)$ of even degree, *every* v-polynomial of a T-(anti)-palindromic pencil in $\mathbb{L}_1(P)$ has at least one root belonging to $\{-1, +1\}$. It follows from the Structured Linearization Theorem that any such $P(\lambda)$ having both $+1$ and -1 as eigenvalues can have neither a T-palindromic nor a T-anti-palindromic linearization in $\mathbb{L}_1(P)$. For T-alternating matrix polynomials $P(\lambda)$ of even degree, every relevant v-polynomial has a root belonging to $\{0, \infty\}$; thus if the spectrum of $P(\lambda)$ includes both 0 and ∞, then P cannot have a T-even or T-odd linearization in $\mathbb{L}_1(P)$.

In situations like the ones above where no structured linearization for $P(\lambda)$ exists in $\mathbb{L}_1(P)$, it is natural to ask whether $P(\lambda)$ has a structured linearization that is *not* in $\mathbb{L}_1(P)$, or perhaps has no structured linearizations at all. The next examples show that either alternative may occur.

Example 7.5.1. Consider the 1×1 T-palindromic polynomial $P(\lambda) = \lambda^2 + 2\lambda + 1$.

The only eigenvalue of $P(\lambda)$ is -1, so by the observation in Example 7.4.2 we see that $P(\lambda)$ cannot have any T-palindromic linearization in $\mathbb{L}_1(P)$. But does $P(\lambda)$ have a T-palindromic linearization $L(\lambda)$ which is not in $\mathbb{L}_1(P)$? Consider the general 2×2 T-palindromic pencil

$$L(\lambda) = \lambda Z + Z^T = \lambda \begin{bmatrix} w & x \\ y & z \end{bmatrix} + \begin{bmatrix} w & y \\ x & z \end{bmatrix}, \tag{7.5.1}$$

and suppose it is a linearization for P. Since the sole eigenvalue of P (i.e. $\lambda = -1$) has geometric multiplicity one, the same must be true for L, that is, $\operatorname{rank} L(-1)$ must be one. But

$$L(-1) = \begin{bmatrix} 0 & y - x \\ x - y & 0 \end{bmatrix}$$

does not have rank one for any values of w, x, y, z. Thus $P(\lambda)$ does not have any T-palindromic linearization at all, either inside or outside of $\mathbb{L}_1(P)$. However, $P(\lambda)$ does have a T-anti-palindromic linearization $\widetilde{L}(\lambda)$ in $\mathbb{L}_1(P)$, because it does not have the eigenvalue $+1$. Choosing $\widetilde{v} = (1, -1)^T$ as right ansatz vector and following the procedure in section 7.3 yields the structured linearization

$$\widetilde{L}(\lambda) = \lambda \widetilde{Z} - \widetilde{Z}^T = \lambda \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix} \in \mathbb{L}_1(P).$$

Example 7.5.2. Consider the T-palindromic matrix polynomial

$$P(\lambda) = \lambda^2 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \lambda^2 A + A^T.$$

Since $\det P(\lambda) = (\lambda^2 - 1)^2$, this polynomial $P(\lambda)$ has $+1$ and -1 as eigenvalues, each with algebraic multiplicity two. Thus $P(\lambda)$ has neither a T-palindromic nor a T-anti-palindromic linearization in $\mathbb{L}_1(P)$. However, it is possible to construct a T-palindromic linearization for $P(\lambda)$ that is *not* in $\mathbb{L}_1(P)$. By a strict equivalence with the first companion linearization $C_1(\lambda) = \lambda \left[\begin{smallmatrix} A & 0 \\ 0 & I \end{smallmatrix}\right] + \left[\begin{smallmatrix} 0 & A^T \\ -I & 0 \end{smallmatrix}\right]$ we obtain another linearization

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot C_1(\lambda) \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \lambda \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

that is clearly T-palindromic. Using shifted sums it can easily be verified that this linearization is in neither $\mathbb{L}_1(P)$ nor $\mathbb{L}_2(P)$. (Of course from the argument in Example 7.4.2 we already know that it could not possibly be in $\mathbb{L}_1(P)$.)

Example 7.5.3. Consider the T-anti-palindromic matrix polynomial $P(\lambda) = \lambda^2 - 1$ with eigenvalues ± 1. Again, the presence of these eigenvalues precludes the existence of either a T-palindromic or T-anti-palindromic linearization in $\mathbb{L}_1(P)$. But even more is true. It turns out that $P(\lambda)$ does not have any T-palindromic or T-anti-palindromic linearization at all. Indeed, suppose that $L_\varepsilon(\lambda) = \lambda Z + \varepsilon Z^T$ was a linearization for $P(\lambda)$ with $\varepsilon = \pm 1$; that is, $L_\varepsilon(\lambda)$ is either T-palindromic or T-anti-palindromic. Since $P(\lambda)$ does not have the eigenvalue ∞, neither does $L(\lambda)$, and so

Z must be invertible. Thus $L_\varepsilon(\lambda)$ is strictly equivalent to the pencil $\lambda I + \varepsilon Z^{-1}Z^T$, which gives us yet another linearization for P. But then the matrix $\varepsilon Z^{-1}Z^T$ would, like $P(\lambda)$, have the simple eigenvalues $+1$ and -1, and hence $\det(\varepsilon Z^{-1}Z^T) = -1$. However, since Z is 2×2 we have

$$\det(\varepsilon Z^{-1}Z^T) = \varepsilon^2 \frac{1}{\det Z} \det Z = +1\,,$$

a contradiction. Thus $P(\lambda)$ has neither a T-palindromic linearization nor a T-anti-palindromic linearization.

These examples clearly show how the presence of eigenvalues $+1$ and -1 may cause trouble in the context of finding structured linearizations for T-palindromic or T-anti-palindromic matrix polynomials. One possibility for circumventing these difficulties is to first deflate the eigenvalues $+1$ and -1 in some kind of structure-preserving manner, using a procedure that works directly on the original matrix polynomial $P(\lambda)$. Since the resulting matrix polynomial $\widehat{P}(\lambda)$ will not have these troublesome eigenvalues, a structured linearization from $\mathbb{L}_1(\widehat{P})$ can then be constructed. Such structure-preserving deflation strategies are currently under investigation.

The situation is quite different for $*$-(anti)-palindromic and $*$-alternating matrix polynomials, because now the set where pairing degenerates is not just $\{+1, -1\}$ or $\{0, \infty\}$, but the entire unit circle in $\mathbb{C}$, or the imaginary axis (including ∞), respectively. The contrast between having a continuum versus a finite set where root pairing degenerates makes a crucial difference in our ability to guarantee the existence of structured linearizations in $\mathbb{L}_1(P)$. Indeed, suppose $P(\lambda)$ is a regular $*$-palindromic matrix polynomial of degree k, and we seek a $*$-palindromic linearization in $\mathbb{L}_1(P)$. Then the v-polynomial $\mathsf{p}(x\,; Rv)$ corresponding to an admissible right ansatz vector v is again $*$-palindromic with $k-1$ roots occurring in pairs $(\lambda, 1/\overline{\lambda}\,)$, by Theorem 6.2.1. Thus if k is even, at least one root of $\mathsf{p}(x\,; Rv)$ must lie on the unit circle. But since the spectrum of $P(\lambda)$ is a finite set, it is always possible to choose v so that all the roots of $\mathsf{p}(x\,; Rv)$ avoid the spectrum of $P(\lambda)$. Here is an illustration for the case $k = 2$.

Example 7.5.4. Consider a regular matrix polynomial $P(\lambda) = \lambda^2 A + \lambda B + A^*$ with $B^* = B$, that is, $P(\lambda)$ is $*$-palindromic. We aim to show that P *always* has a structured linearization in $\mathbb{L}_1(P)$; the main problem is to decide how to choose a suitable right ansatz vector. Since P has only finitely many eigenvalues, there is some $\zeta \in \mathbb{C}$ of unit modulus such that ζ is not an eigenvalue of $P(\lambda)$; let $\alpha \in \mathbb{C}$ be such that $\zeta = -\alpha/\overline{\alpha}$. From Theorem 7.2.1 we know that the right ansatz vector of any $*$-palindromic pencil in $\mathbb{L}_1(P)$ must satisfy $Rv = \overline{v}$. Thus $v = [\,\alpha, \overline{\alpha}\,]^T$ is an admissible right ansatz vector, and the associated v-polynomial $\mathsf{p}(x\,; Rv) = \overline{\alpha}x + \alpha$ has only the root ζ. Therefore by Theorem 7.4.1 the pencil

$$L(\lambda) = \lambda \begin{bmatrix} \alpha A & \alpha B - \overline{\alpha} A^* \\ \overline{\alpha} A & \alpha A \end{bmatrix} + \begin{bmatrix} \overline{\alpha} A^* & \alpha A^* \\ \overline{\alpha} B - \alpha A & \overline{\alpha} A^* \end{bmatrix} \in \mathbb{L}_1(P)$$

with right ansatz vector v is a $*$-palindromic linearization for $P(\lambda)$.

It should be noted that all the observations made in this section for $\star$-(anti)-palindromic polynomials have parallels for the case of $\star$-alternating structures. See

Tables 7.7.1, 7.7.2, and 7.7.3 for a list of structured pencils from $\mathbb{L}_1(P)$ for $\star$-(anti)-palindromic and $\star$-alternating matrix polynomials of degree $k = 2, 3$, together with their corresponding "eigenvalue avoidance" linearization conditions.

7.6 Good Vibrations from Good Linearizations

As an illustration of the importance of structure preservation in practical problems we reconsider Example 6.4.1, and briefly indicate how the techniques developed in this thesis have had a significant impact on computations in this application.

Recall that the eigenvalue problem in this example comes from the vibration analysis of rail tracks under excitation from high speed trains, and has the form

$$\left(\kappa A(\omega) + B(\omega) + \frac{1}{\kappa} A(\omega)^T\right) x = 0 \tag{7.6.1}$$

where A, B are large, sparse, parameter-dependent, complex square matrices with B complex symmetric and A highly singular. For details of the derivation of this model see [37] and [38]. The parameter ω is the excitation frequency and the eigenvalue problem has to be solved over a wide frequency range of $\omega =$ 0-5,000 Hz. Clearly, for any fixed value of ω, multiplying (7.6.1) by κ leads to the T-palindromic eigenvalue problem introduced in (6.1.1). In addition to the presence of a large number of zero and infinite eigenvalues caused by the rank deficiency of A, the finite nonzero eigenvalues cover a wide range of magnitudes that increases as the finite element discretization is made finer. The eigenvalues of a typical example of this problem range from 10^{-15}–10^{15}, thereby making this a very challenging numerical problem.

Attempts at solving this problem with the QZ-algorithm without respecting its structure resulted in computed eigenvalues with no correct digits even in quadruple precision arithmetic. Furthermore, the symmetry of the spectrum with respect to the unit circle was highly perturbed [37].

As an alternative, in [37], [38] a T-palindromic linearization for (7.6.1) was used. Based on this linearization the infinite and zero eigenvalues of the resulting T-palindromic pencil could be deflated in a structure-preserving way. The resulting smaller T-palindromic problem was then solved via different methods, resulting in eigenvalues with good accuracy in double precision arithmetic, i.e., within the range of the discretization error of the underlying finite element discretization. Thus physically useful eigenvalues were determined *without any change* in either the mathematical model or the discretization scheme. The only change made was in the numerical linear algebra, to methods based on the new structured linearization techniques in this thesis.

Thus we see that the computation of "good vibrations" (i.e., accurate eigenvalues and eigenvectors) requires the use of "good linearizations" (i.e., linearizations that reflect the structure of the original polynomial).

7.7 Structured Subspaces of $\mathbb{L}_1(P)$

Although they have not been given their own name or notation, or had their properties explored in any detail, this chapter should not end without at least briefly mentioning the various *structured subspaces* of the pencil space $\mathbb{L}_1(P)$ and of the corresponding ansatz vector space $\mathbb{F}^k$ that are "implicitly" defined by the results in section 7.2. These subspaces can be viewed as the analogs for palindromic and alternating polynomials of the structured subspaces $\mathbb{S}(P)$ and $\mathbb{H}(P)$ for symmetric and Hermitian polynomials P that were discussed in chapter 5.

Suppose, for example, that $P(\lambda)$ is T-palindromic of degree k. Then Theorem 7.2.1 shows that in $\mathbb{L}_1(P)$ the sets of all T-palindromic and T-anti-palindromic pencils form two nontrivial structured subspaces, of dimension $\lceil k/2 \rceil$ and $\lfloor k/2 \rfloor$, respectively; their corresponding admissible right ansatz vector sets also constitute structured subspaces of $\mathbb{F}^k$ with the same dimensions $\lceil k/2 \rceil$ and $\lfloor k/2 \rfloor$. Furthermore these two ansatz vector subspaces are simply related to each other; they are orthogonal complements in $\mathbb{F}^k$. Similar results can also be shown to hold for the other seven types of structured matrix polynomial considered in Theorems 7.2.1 and 7.2.2.

Table 7.7.1: Structured linearizations for $P(\lambda) = \lambda^2 A + \lambda B + C$. Except for the parameters $r \in \mathbb{R}$ and $z \in \mathbb{C}$, the linearizations are unique up to a (suitable) scalar factor. The last column lists the roots of the v-polynomial $\mathsf{p}(x\,; w)$ corresponding to $w = Rv$ (for palindromic structures) or $w = \Sigma v$ (for alternating structures); for L to be a linearization these eigenvalues must be excluded from P.

Structure of $P(\lambda)$	Structure of $L(\lambda)$	v	$L(\lambda)$ with right ansatz vector v	Root of $\mathsf{p}(x\,;w)$
T-palindromic	T-palindromic	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} A & B-C \\ A & A \end{bmatrix} + \begin{bmatrix} C & C \\ B-A & C \end{bmatrix}$	-1
$B = B^T$, $C = A^T$	T-anti-palindromic	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\lambda \begin{bmatrix} A & B+C \\ -A & A \end{bmatrix} + \begin{bmatrix} -C & C \\ -B-A & -C \end{bmatrix}$	1
T-anti-palindromic.	T-palindromic	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\lambda \begin{bmatrix} A & B+C \\ -A & A \end{bmatrix} + \begin{bmatrix} -C & C \\ -B-A & -C \end{bmatrix}$	1
$B = -B^T$, $C = -A^T$	T-anti-palindromic	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} A & B-C \\ A & A \end{bmatrix} + \begin{bmatrix} C & C \\ B-A & C \end{bmatrix}$	-1
$*$-palindromic	$*$-palindromic	$\begin{bmatrix} z \\ \bar{z} \end{bmatrix}$	$\lambda \begin{bmatrix} zA & zB-\bar{z}C \\ \bar{z}A & zA \end{bmatrix} + \begin{bmatrix} \bar{z}C & zC \\ \bar{z}B-zA & \bar{z}C \end{bmatrix}$	$-z/\bar{z}$
$B = B^*$, $C = A^*$	$*$-anti-palindromic	$\begin{bmatrix} z \\ -\bar{z} \end{bmatrix}$	$\lambda \begin{bmatrix} zA & zB+\bar{z}C \\ -\bar{z}A & zA \end{bmatrix} + \begin{bmatrix} -\bar{z}C & zC \\ -\bar{z}B-zA & -\bar{z}C \end{bmatrix}$	$z/\bar{z}$
T-even	T-even	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} 0 & -A \\ A & B \end{bmatrix} + \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$	∞
$A = A^T$, $B = -B^T$, $C = C^T$	T-odd	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\lambda \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix} + \begin{bmatrix} B & C \\ -C & 0 \end{bmatrix}$	0
T-odd	T-even	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\lambda \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix} + \begin{bmatrix} B & C \\ -C & 0 \end{bmatrix}$	0
$A = -A^T$, $B = B^T$, $C = -C^T$	T-odd	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} 0 & -A \\ A & B \end{bmatrix} + \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$	∞
$*$-even	$*$-even	$\begin{bmatrix} i \\ r \end{bmatrix}$	$\lambda \begin{bmatrix} iA & -rA \\ rA & rB+iC \end{bmatrix} + \begin{bmatrix} rA+iB & iC \\ -iC & rC \end{bmatrix}$	$-ir$
$A = A^*$, $B = -B^*$, $C = C^*$	$*$-odd	$\begin{bmatrix} r \\ i \end{bmatrix}$	$\lambda \begin{bmatrix} rA & -iA \\ iA & iB+rC \end{bmatrix} + \begin{bmatrix} iA+rB & rC \\ -rC & iC \end{bmatrix}$	$\frac{i}{r}$

Table 7.7.2: $\star$-palindromic linearizations for the $\star$-palindromic matrix polynomial $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda B^\star + A^\star$. The last column lists the roots of the v-polynomial corresponding to Rv. All $\star$-palindromic linearizations in $\mathbb{L}_1(P)$ for this matrix polynomial are linear combinations of the first two linearizations in the case $\star = T$, and real linear combinations of the first three linearizations in the case $\star = *$. A specific example is given by the fourth linearization.

v	$L(\lambda)$ with right ansatz vector v	Roots of $\mathsf{p}(x\,; Rv)$
$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	$\lambda \begin{bmatrix} 0 & 0 & -A^\star \\ A & B & 0 \\ 0 & A & 0 \end{bmatrix} + \begin{bmatrix} 0 & A^\star & 0 \\ 0 & B^\star & A^\star \\ -A & 0 & 0 \end{bmatrix}$	$0, \infty$
$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} A & B - A^\star & B^\star \\ 0 & A - B^\star & B - A^\star \\ A & 0 & A \end{bmatrix} + \begin{bmatrix} A^\star & 0 & A^\star \\ B^\star - A & A^\star - B & 0 \\ B & B^\star - A & A^\star \end{bmatrix}$	$i, -i$
$\begin{bmatrix} i \\ 0 \\ -i \end{bmatrix}$	$\lambda \begin{bmatrix} iA & iB + iA^* & iB^* \\ 0 & iA + iB^* & iB + iA^* \\ -iA & 0 & iA \end{bmatrix} + \begin{bmatrix} -iA^* & 0 & iA^* \\ -iB^* - iA & -iA^* - iB & 0 \\ -iB & -iB^* - iA & -iA^* \end{bmatrix}$	$1, -1$
$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} A & B - A^\star & B^\star - A^\star \\ A & B + A - B^\star & B - A^\star \\ A & A & A \end{bmatrix} + \begin{bmatrix} A^\star & A^\star & A^\star \\ B^\star - A & B^\star + A^\star - B & A^\star \\ B - A & B^\star - A & A^\star \end{bmatrix}$	$\dfrac{-1 \pm i\sqrt{3}}{2}$

Table 7.7.3: $\star$-even linearizations for the $\star$-even matrix polynomial $P(\lambda) = \lambda^3 A + \lambda^2 B + \lambda C + D$, where $A = -A^\star$, $B = B^\star$, $C = -C^\star$, $D = D^\star$. The last column lists the roots of the v-polynomial corresponding to Σv. All $\star$-even linearizations in $\mathbb{L}_1(P)$ for this matrix polynomial are linear combinations of the first two linearizations in the case $\star = T$, and real linear combinations of the first three linearizations in the case $\star = *$. A specific example is given by the fourth linearization.

v	$L(\lambda)$ with right ansatz vector v	Roots of $\mathsf{p}(x\,; \Sigma v)$
$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\lambda \begin{bmatrix} 0 & 0 & A \\ 0 & -A & -B \\ A & B & C \end{bmatrix} + \begin{bmatrix} 0 & -A & 0 \\ A & B & 0 \\ 0 & 0 & D \end{bmatrix}$	∞
$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$	$\lambda \begin{bmatrix} A & 0 & 0 \\ 0 & C & D \\ 0 & -D & 0 \end{bmatrix} + \begin{bmatrix} B & C & D \\ -C & -D & 0 \\ D & 0 & 0 \end{bmatrix}$	0
$\begin{bmatrix} 0 \\ i \\ 0 \end{bmatrix}$	$\lambda \begin{bmatrix} 0 & -iA & 0 \\ iA & iB & 0 \\ 0 & 0 & iD \end{bmatrix} + \begin{bmatrix} iA & 0 & 0 \\ 0 & iC & iD \\ 0 & -iD & 0 \end{bmatrix}$	$0, \infty$
$\begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}$	$\lambda \begin{bmatrix} A & 0 & 4A \\ 0 & C - 4A & D - 4B \\ 4A & 4B - D & 4C \end{bmatrix} + \begin{bmatrix} B & C - 4A & D \\ 4A - C & 4B - D & 0 \\ D & 0 & 4D \end{bmatrix}$	$2i, -2i$

Chapter 8

Conditioning of Eigenvalues of Linearizations

In this thesis we have developed two pencil spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ associated to any matrix polynomial P, and shown that they provide large sources of linearizations for P. The large size of these spaces has enabled us to find subspaces like $\mathbb{B}(P)$ and $\mathbb{DL}(P)$ of pencils with special properties, and to find linearizations that reflect additional structures of P. But the very size of these spaces poses another problem: among all these choices of linearization, how do you pick which one to compute with? Is there a "best" linearization among the infinitely many possible pencils in these spaces?

In this final chapter we begin to address these questions by considering the sensitivity of eigenvalues of linearizations, with the aim of providing at least some guidance on how to make this choice. The focus is on pencils in $\mathbb{DL}(P)$ because of their key role in finding structured linearizations, and because the special properties of these pencils make for a relatively clean analysis. We also give a separate treatment of two particular linearizations that are not in $\mathbb{DL}(P)$, the two companion linearizations C_1 and C_2 (see Definition 1.1.5), because of their frequent use in current computational practice. The results of this analysis gives some insight into potential instability of the companion linearizations. Finally, some numerical experiments in Section 8.7 illustrate the ability of our analysis to predict well the accuracy of eigenvalues computed via different linearizations.

8.1 Eigenvalue Conditioning of Matrix Polynomials

Let λ be a simple, finite, nonzero eigenvalue of $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ with corresponding right eigenvector x and left eigenvector y. A normwise condition number of λ can be defined by

$$\kappa_P(\lambda) = \lim_{\epsilon \to 0} \sup \left\{ \frac{|\Delta\lambda|}{\epsilon|\lambda|} : \big(P(\lambda + \Delta\lambda) + \Delta P(\lambda + \Delta\lambda)\big)(x + \Delta x) = 0, \right.$$
$$\left. \|\Delta A_i\|_2 \le \epsilon\omega_i,\ i = 0\colon k \right\}, \qquad (8.1.1)$$

where $\Delta P(\lambda) = \sum_{i=0}^{k} \lambda^i \Delta A_i$. The ω_i are nonnegative weights that allow flexibility in how the perturbations are measured; in particular, ΔA_i can be forced to zero by setting $\omega_i = 0$. An explicit formula for this condition number is given in the following result.

Theorem 8.1.1 (Tisseur [80, Thm. 5]). *The normwise condition number $\kappa_P(\lambda)$ is given by*

$$\kappa_P(\lambda) = \frac{\left(\sum_{i=0}^{k} |\lambda|^i \omega_i\right) \|y\|_2 \|x\|_2}{|\lambda|\, |y^* P'(\lambda) x|}. \tag{8.1.2}$$

The condition number $\kappa_P(\lambda)$ has the unfortunate disadvantage that it is not defined for zero or infinite eigenvalues. In order to give a unified treatment for all λ, we rewrite the polynomial in the homogeneous form

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

and consider eigenvalues as pairs $(\alpha, \beta) \neq (0,0)$ that are solutions of the scalar equation $\det P(\alpha, \beta) = 0$; here $\lambda \equiv \alpha/\beta$. More precisely, since $P(\alpha, \beta)$ is homogeneous in α and β, we define an eigenvalue as any line through the origin in $\mathbb{C}^2 \setminus \{0\}$ of solutions of $\det P(\alpha, \beta) = 0$. Let $T_{(\alpha,\beta)}\mathbb{P}_1$ denote the tangent space at (α, β) to $\mathbb{P}_1$, the projective space of lines through the origin in $\mathbb{C}^2 \setminus \{0\}$. Dedieu and Tisseur [18] define a *condition operator* $K(\alpha, \beta) : (\mathbb{C}^{n\times n})^{k+1} \to T_{(\alpha,\beta)}\mathbb{P}_1$ for the eigenvalue (α, β) as the differential of the map from the $(k+1)$-tuple $(A_0, \dots, A_k)$ to (α, β) in projective space. The significance of this condition operator is shown by the following result, which is an extension of a result of Dedieu [17, Thm. 6.1]. Here and below, we sometimes write a representative of an eigenvalue (α, β) as a nonzero row vector $[\,\alpha, \beta\,] \in \mathbb{C}^{1\times 2}$.

Theorem 8.1.2. *Let (α, β) be a simple eigenvalue of $P(\alpha, \beta)$ with representative $[\,\alpha, \beta\,]$ normalized so that $\|[\,\alpha, \beta\,]\|_2 = 1$. For sufficiently small $(k+1)$-tuples*

$$\Delta A \equiv (\Delta A_0, \dots, \Delta A_k),$$

the perturbed polynomial $\widetilde{P}(\alpha, \beta) = \sum_{i=0}^{k} \alpha^i \beta^{k-i}(A_i + \Delta A_i)$ has a simple eigenvalue $(\widetilde{\alpha}, \widetilde{\beta})$ for which, with the normalization $[\,\alpha, \beta\,][\,\widetilde{\alpha}, \widetilde{\beta}\,]^* = 1$,

$$[\,\widetilde{\alpha}, \widetilde{\beta}\,] = [\,\alpha, \beta\,] + K(\alpha, \beta)\Delta A + o(\|\Delta A\|).$$

A "homogeneous condition number " $\widehat{\kappa}_P(\alpha, \beta)$ can now be defined as a norm of the condition operator:

$$\widehat{\kappa}_P(\alpha, \beta) = \max_{\|\Delta A\| \le 1} \frac{\|K(\alpha, \beta)\Delta A\|_2}{\|\,[\,\alpha, \beta\,]\,\|_2},$$

where the norm on ΔA is arbitrary. Note that this condition number is well-defined, since the right-hand side is independent of the choice of representative of the eigenvalue (α, β). Let $\theta\big((\mu, \nu), (\widetilde{\mu}, \widetilde{\nu})\big)$ be the angle between the two lines (μ, ν) and $(\widetilde{\mu}, \widetilde{\nu})$. Then for θ small enough,

$$|\theta\big((\mu, \nu), (\widetilde{\mu}, \widetilde{\nu})\big)| \le \big|\tan\big(\theta\big((\mu, \nu), (\widetilde{\mu}, \widetilde{\nu})\big)\big)\big| = \left\| [\,\widetilde{\mu}, \widetilde{\nu}\,] \frac{\|[\,\mu, \nu\,]\|_2}{[\,\widetilde{\mu}, \widetilde{\nu}\,][\,\mu, \nu\,]^*} - \frac{[\,\mu, \nu\,]}{\|[\,\mu, \nu\,]\|_2} \right\|_2.$$

Inserting the particular representatives $[\,\alpha,\beta\,]$ and $[\,\widetilde{\alpha},\widetilde{\beta}\,]$ of the original and perturbed eigenvalues, normalized as in Theorem 8.1.2, gives

$$\big|\theta\big((\alpha,\beta),(\widetilde{\alpha},\widetilde{\beta})\big)\big| \le \|\,[\,\alpha,\beta\,]-[\,\widetilde{\alpha},\widetilde{\beta}\,]\,\|_2 = \|K(\alpha,\beta)\Delta A\|_2 + o(\|\Delta A\|).$$

Hence, the angle between the original and perturbed eigenvalues satisfies

$$\big|\theta\big((\alpha,\beta),(\widetilde{\alpha},\widetilde{\beta})\big)\big| \le \widehat{\kappa}_P(\alpha,\beta)\|\Delta A\| + o(\|\Delta A\|). \tag{8.1.3}$$

By taking the sine of both sides we obtain a perturbation bound in terms of $\sin|\theta|$, which is the chordal distance between (α,β) and $(\widetilde{\alpha},\widetilde{\beta})$ as used by Stewart and Sun [79, Chap. 6]. Of course, $\sin|\theta| \le |\theta|$ and asymptotically these two measures of distance are equal.

We will take for the norm on $(\mathbb{C}^{n\times n})^{k+1}$ the ω-weighted Frobenius norm

$$\|A\| = \|(A_0,\ldots,A_k)\| = \|[\,\omega_0^{-1}A_0,\ldots,\omega_k^{-1}A_k\,]\|_F, \tag{8.1.4}$$

where the ω_i are nonnegative weights that are analogous to those in (8.1.1). Define the operators $\mathcal{D}_\alpha \equiv \frac{\partial}{\partial\alpha}$ and $\mathcal{D}_\beta \equiv \frac{\partial}{\partial\beta}$. Then the following theorem is a trivial extension of a result of Dedieu and Tisseur [18, Thm. 4.2] that treats the unweighted Frobenius norm.

Theorem 8.1.3. *The normwise condition number $\widehat{\kappa}_P(\alpha,\beta)$ of a simple eigenvalue (α,β) is given by*

$$\widehat{\kappa}_P(\alpha,\beta) = \left(\sum_{i=0}^{k}|\alpha|^{2i}|\beta|^{2(k-i)}\omega_i^2\right)^{1/2}\frac{\|y\|_2\,\|x\|_2}{\big|y^*(\bar{\beta}\mathcal{D}_\alpha P - \bar{\alpha}\mathcal{D}_\beta P)|_{(\alpha,\beta)}x\big|}. \tag{8.1.5}$$

As a check, we note that the expression (8.1.5) is independent of the choice of representative of (α,β) and of the scaling of x and y. Note also that for a simple eigenvalue the denominator terms $y^*P'(\lambda)x$ in (8.1.2) and $y^*(\bar{\beta}\mathcal{D}_\alpha - \bar{\alpha}\mathcal{D}_\beta)P|_{(\alpha,\beta)}x$ in (8.1.5) are both nonzero, as shown in [2, Thm. 3.2] for the former and [18, Thm. 3.3 (iii)] for the latter.

To summarize, the condition numbers $\kappa_P(\lambda)$ and $\widehat{\kappa}_P(\alpha,\beta)$ give two different measures of the sensitivity of a simple eigenvalue. The advantage of $\kappa_P(\lambda)$ is that it is an immediate generalization of the well-known Wilkinson condition number for the standard eigenproblem [86, p. 69] and it measures the relative change in an eigenvalue, which is a concept readily understood by users of numerical methods. In favor of $\widehat{\kappa}_P(\alpha,\beta)$ is that it elegantly treats all eigenvalues, including those at zero and infinity; moreover, it provides the bound (8.1.3) for the angular error, which is an alternative to the relative error bound that $\kappa_P(\lambda)$ provides. Both condition numbers are therefore of interest, and hence will both be treated throughout the rest of this chapter.

We note that in MATLAB 7.0 (R14) the function `polyeig` that solves the polynomial eigenvalue problem returns the condition number $\widehat{\kappa}_P(\alpha,\beta)$ as an optional output argument.

8.2 Eigenvalue Conditioning of $\mathbb{DL}(P)$-Linearizations

We now focus on the condition numbers $\kappa_L(\lambda)$ and $\widehat{\kappa}_L(\alpha,\beta)$ of a simple eigenvalue of a linearization $L(\lambda) = \lambda X + Y$ from $\mathbb{DL}(P)$. Our aim is to obtain expressions for these condition numbers that have two properties: they should separate the dependence on P from that of the ansatz vector v defining L, and they should have minimal explicit dependence on X and Y. In the next section we will then consider how to minimize these expressions over all v. Note the distinction between the condition numbers κ_L and $\widehat{\kappa}_L$ of the pencil and κ_P and $\widehat{\kappa}_P$ of the original polynomial. Note also that a simple eigenvalue of L is necessarily a simple eigenvalue of P, and vice versa, in view of Definition 1.1.4.

We first carry out the analysis for $\widehat{\kappa}_L(\alpha,\beta)$. Let x and y denote right and left eigenvectors of P, and z and w denote right and left eigenvectors of L, all corresponding to the eigenvalue (α,β). Recalling that $\lambda = \alpha/\beta$, define

$$\begin{aligned} L(\alpha,\beta) \;&:=\; \alpha X + \beta Y \;=\; \beta L(\lambda) \\ \text{and}\quad \Lambda_{\alpha,\beta} \;&:=\; [\alpha^{k-1}, \alpha^{k-2}\beta, \ldots, \beta^{k-1}]^T \;=\; \beta^{k-1}\Lambda\,. \end{aligned} \tag{8.2.1}$$

In view of the relations in Theorems 2.1.8, 2.1.13, and 2.2.4 between eigenvectors of P and those of L we can take

$$w = \overline{\Lambda}_{\alpha,\beta} \otimes y, \quad z = \Lambda_{\alpha,\beta} \otimes x. \tag{8.2.2}$$

(These expressions are valid for both finite and infinite eigenvalues.) The condition number $\widehat{\kappa}_L(\alpha,\beta)$ can now be evaluated by applying Theorem 8.1.3 to L:

$$\widehat{\kappa}_L(\alpha,\beta) = \sqrt{|\alpha|^2\omega_X^2 + |\beta|^2\omega_Y^2}\;\frac{\|w\|_2\,\|z\|_2}{\left|w^*(\bar{\beta}\,\mathcal{D}_\alpha L - \bar{\alpha}\,\mathcal{D}_\beta L)|_{(\alpha,\beta)} z\right|}\,, \tag{8.2.3}$$

where an obvious notation has been used for the weights in (8.1.4).

In the homogeneous notation (8.2.1) the condition in Definition (2.1.1) that characterizes a member of $\mathbb{L}_1(P)$ can be rewritten as

$$L(\alpha,\beta)(\Lambda_{\alpha,\beta} \otimes I_n) = v \otimes P(\alpha,\beta)\,, \tag{8.2.4}$$

where for the moment α and β denote variables. Differentiating with respect to α gives

$$\mathcal{D}_\alpha L(\alpha,\beta)(\Lambda_{\alpha,\beta} \otimes I_n) + L(\alpha,\beta)(\mathcal{D}_\alpha \Lambda_{\alpha,\beta} \otimes I_n) = v \otimes \mathcal{D}_\alpha P(\alpha,\beta). \tag{8.2.5}$$

Now evaluate this equation at an eigenvaluefootnoteStrictly speaking, here and later we are evaluating at a representative of an eigenvalue. All the condition number formulae are independent of the choice of representative. (α,β). Multiplying on the left by w^* and on the right by $1 \otimes x$, and using (8.2.2), we obtain

$$\begin{aligned} w^*(\mathcal{D}_\alpha L)|_{(\alpha,\beta)} z \;&=\; \Lambda_{\alpha,\beta}^T v \otimes y^*(\mathcal{D}_\alpha P)|_{(\alpha,\beta)} x \\ &=\; \Lambda_{\alpha,\beta}^T v \cdot y^*(\mathcal{D}_\alpha P)|_{(\alpha,\beta)} x. \end{aligned} \tag{8.2.6}$$

Exactly the same argument leads to

$$w^*(\mathcal{D}_\beta L)|_{(\alpha,\beta)} z = \Lambda_{\alpha,\beta}^T v \cdot y^*(\mathcal{D}_\beta P)|_{(\alpha,\beta)} x. \tag{8.2.7}$$

Hence, from (8.2.6) and (8.2.7),

$$w^*(\bar\beta\mathcal{D}_\alpha L - \bar\alpha\mathcal{D}_\beta L)|_{(\alpha,\beta)}z = \Lambda_{\alpha,\beta}^T v \cdot y^*(\bar\beta\mathcal{D}_\alpha P - \bar\alpha\mathcal{D}_\beta P)|_{(\alpha,\beta)}x.$$

The first factor on the right can be viewed as the homogeneous version

$$\mathsf{p}(\alpha,\beta\,;v) \;:=\; \Lambda_{\alpha,\beta}^T v \;=\; v^T\Lambda_{\alpha,\beta} \;=\; \sum_{i=1}^{k} v_i\alpha^{k-i}\beta^{i-1} \;=\; \beta^k\mathsf{p}(\lambda\,;v) \tag{8.2.8}$$

of the scalar v-polynomial $\mathsf{p}(\lambda\,;v) = \Lambda^T v$ introduced in Definition 4.2.1.

Noting, from (8.2.2), that $\|w\|_2 = \|\Lambda_{\alpha,\beta}\|_2\,\|y\|_2$ and $\|z\|_2 = \|\Lambda_{\alpha,\beta}\|_2\,\|x\|_2$, we obtain an alternative form of (8.2.3) that clearly separates the dependence of $\widehat\kappa_L$ on P from its dependence on the ansatz vector v. From now on the extended notation $\widehat\kappa_L(\alpha,\beta\,;v)$ will be used to emphasize this dependence of $\widehat\kappa_L$ on the vector $v\in\mathbb{C}^k$ that defines the linearization in $\mathbb{DL}(P)$.

Theorem 8.2.1. *Let (α,β) be a simple eigenvalue of P with right and left eigenvectors x and y, respectively. Then for any pencil $L(\alpha,\beta)=\alpha X+\beta Y\in\mathbb{DL}(P)$ that is a linearization of P,*

$$\widehat\kappa_L(\alpha,\beta\,;v) = \frac{\sqrt{|\alpha|^2\omega_X^2+|\beta|^2\omega_Y^2}}{|\mathsf{p}(\alpha,\beta\,;v)|}\cdot\frac{\|\Lambda_{\alpha,\beta}\|_2^2\,\|y\|_2\,\|x\|_2}{\left|\,y^*(\bar\beta\mathcal{D}_\alpha P-\bar\alpha\mathcal{D}_\beta P)|_{(\alpha,\beta)}x\,\right|}\,, \tag{8.2.9}$$

where v is the ansatz vector of $L(\alpha,\beta)$ as in (8.2.4).

Now we give a similar analysis for the condition number $\kappa_L(\lambda)$ of a simple, finite, nonzero eigenvalue λ. In view of (8.1.2), our aim is to obtain an expression for $|w^*L'(\lambda)z|$. Since $L\in\mathbb{L}_1(P)$,

$$L(\lambda)(\Lambda\otimes I_n) = v\otimes P(\lambda). \tag{8.2.10}$$

Differentiating the identity (8.2.10) with respect to λ gives

$$L'(\lambda)(\Lambda\otimes I_n) + L(\lambda)(\Lambda'\otimes I_n) = v\otimes P'(\lambda). \tag{8.2.11}$$

Evaluating at an eigenvalue λ, premultiplying by $w^*=\Lambda^T\otimes y^*$, postmultiplying by $1\otimes x$, and using (8.2.2) gives

$$w^*L'(\lambda)z = \Lambda^T v\otimes y^*P'(\lambda)x = \mathsf{p}(\lambda\,;v)\cdot y^*P'(\lambda)x\,,$$

and thus the following analog of Theorem 8.2.1.

Theorem 8.2.2. *Let λ be a simple, finite, nonzero eigenvalue of P with right and left eigenvectors x and y, respectively. Then for any pencil $L(\lambda)=\lambda X+Y\in\mathbb{DL}(P)$ that is a linearization of P,*

$$\kappa_L(\lambda\,;v) = \frac{(|\lambda|\omega_X+\omega_Y)}{|\mathsf{p}(\lambda\,;v)|}\cdot\frac{\|\Lambda\|_2^2\,\|y\|_2\,\|x\|_2}{|\lambda|\;|y^*P'(\lambda)x|}\,, \tag{8.2.12}$$

where v is the ansatz vector of $L(\lambda)$ as in (8.2.10).

The expression (8.2.9) shows that $\widehat\kappa_L(\alpha,\beta)$ is finite if and only if (α,β) is not a zero of $\mathsf{p}(\alpha,\beta\,;v)$, and (8.2.12) gives essentially the same information for $\lambda\neq 0,\infty$. This result dovetails nicely with the Eigenvalue Exclusion Theorem from section 4.3, which shows that $L(\lambda)$ is a linearization for $P(\lambda)$ if and only if no eigenvalue of P (including ∞) is a root of $\mathsf{p}(\lambda\,;v)$.

8.3 Minimizing $\widehat{\kappa}_L(\alpha,\beta)$ and $\kappa_L(\lambda)$

Recall that pencils in $\mathbb{DL}(P)$ are uniquely defined by their ansatz vectors v. Our aim in this section is to minimize the condition numbers $\kappa_L(\lambda; v)$ and $\widehat{\kappa}_L(\alpha,\beta\,; v)$ over all $v \in \mathbb{C}^k$, thereby identifying a best conditioned linearization in $\mathbb{DL}(P)$ for a particular eigenvalue.

A technical subtlety here is that the minimum of $\kappa_L(\lambda; v)$ or $\widehat{\kappa}_L(\alpha,\beta\,; v)$ over v could potentially occur at an ansatz vector v for which $L(\lambda)$ is not a linearization; note that the formulas (8.2.9) and (8.2.12) depend only on a particular eigenvalue, whereas Theorem 4.3.1 says that the property of being a linearization involves *all* the eigenvalues. In this case formulas (8.2.9) and (8.2.12) are not valid. However, such "bad" v form a closed, nowhere dense set of measure zero [Thm. 4.3.2], so an arbitrarily small perturbation to v can make L a linearization.

Expressions (8.2.9) and (8.2.12) have similar forms, with dependence on v confined to the $\mathsf{p}(\cdot)$ terms in the denominator and the ω terms in the numerator. For most of this section we work with the condition number (8.2.9) for the pencil in homogeneous form, returning to $\kappa_L(\lambda)$ at the end of the section.

For the weights we will take the natural choice

$$\omega_X = \|X\|_2, \quad \omega_Y = \|Y\|_2. \tag{8.3.1}$$

Since by Theorems 3.3.2 and 3.4.2 the entries of X and Y are linear combinations of the entries of v, this choice of weights makes the condition numbers independent of the scaling of v.

We consider first the v-dependence of $\|X\|_2$ and $\|Y\|_2$.

Lemma 8.3.1. *For $L(\lambda) = \lambda X + Y \in \mathbb{DL}(P)$ defined by $v \in \mathbb{C}^k$ we have*

$$\|v\|_2\,\|A_k\|_2 \;\le\; \|X\|_2 \;\le\; k\,r^{1/2}\max_i \|A_i\|_2\,\|v\|_2, \tag{8.3.2}$$

$$\|v\|_2\,\|A_0\|_2 \;\le\; \|Y\|_2 \;\le\; k\,r^{1/2}\max_i \|A_i\|_2\,\|v\|_2, \tag{8.3.3}$$

where r is the number of nonzero entries in v.

Proof. Partition X and Y as block $k\times k$ matrices with $n\times n$ blocks. From Theorem 2.1.5 we know that the first block column of X is $v\otimes A_k$ and the last block column of Y is $v\otimes A_0$. The lower bounds are therefore immediate. From the form of the standard basis for $\mathbb{DL}(P)$ in Theorem 3.3.2 it can be seen that each block X_{pq} of X can be expressed as a sum of the form

$$X_{pq} = \sum_{i=1}^{k} s_i v_i A_{\ell_i}, \tag{8.3.4}$$

where $s_i \in \{-1,0,1\}$ and the indices ℓ_i are *distinct*. Hence

$$\|X_{pq}\|_2 \le \max_i \|A_i\|_2 \sum_{i=1}^{k} |v_i| = \max_i \|A_i\|_2\,\|v\|_1 \le r^{1/2}\max_i \|A_i\|_2\,\|v\|_2\,.$$

The upper bound on $\|X\|_2$ follows on using

$$\|X\|_2 \le k\max_{p,q}\|X_{pq}\|_2\,, \tag{8.3.5}$$

which holds for any block $k \times k$ matrix. An identical argument gives the upper bound for $\|Y\|_2$. ∎

Hence, provided the $\|A_i\|_2$ values vary little in magnitude with i, the numerator of (8.2.9) varies little in magnitude with v if $\|v\|_2$ is fixed. Under this proviso, we will approximately minimize the condition number $\widehat{\kappa}_L(\alpha,\beta)$ if we maximize the $\mathsf{p}(\alpha,\beta\,;v)$ term. We therefore restrict our attention to the denominator of the expression (8.2.9) for $\widehat{\kappa}_L$ and maximize $|\mathsf{p}(\alpha,\beta\,;v)| = |\Lambda_{\alpha,\beta}^T v|$ subject to $\|v\|_2 = 1$, for a given eigenvalue (α,β). By the Cauchy–Schwarz inequality the maximizing v and the corresponding value of the polynomial are

$$v_* = \frac{\overline{\Lambda}_{\alpha,\beta}}{\|\Lambda_{\alpha,\beta}\|_2}, \qquad |\mathsf{p}(\alpha,\beta\,;v_*)| = \|\Lambda_{\alpha,\beta}\|_2. \tag{8.3.6}$$

Two special cases that play an important role in the rest of this paper are worth noting:

$$\begin{aligned} (\alpha,\beta) = (1,0), \quad \lambda = \infty \quad &\Rightarrow \quad v_* = e_1,\\ (\alpha,\beta) = (0,1), \quad \lambda = 0 \quad &\Rightarrow \quad v_* = e_k. \end{aligned}$$

The next theorem compares the condition numbers for $v = e_1$ and $v = e_k$ with the optimal condition numberamong all linearizations in $\mathbb{DL}(P)$. Define

$$\rho = \frac{\max_i \|A_i\|_2}{\min(\|A_0\|_2, \|A_k\|_2)} \geq 1. \tag{8.3.7}$$

When we write $\inf_v \widehat{\kappa}_L(\alpha,\beta\,;v)$ the infimum is understood to be taken over v for which L is a linearization.

Theorem 8.3.2. *Let (α,β) be a simple eigenvalue of P and consider pencils $L \in \mathbb{DL}(P)$. Take the weights* (8.3.1) *for $\widehat{\kappa}_L$. Then*

$$\widehat{\kappa}_L(\alpha,\beta\,;e_1) \;\leq\; \rho k^{3/2} \inf_v \widehat{\kappa}_L(\alpha,\beta\,;v) \;\; \textit{if } A_0 \textit{ is nonsingular and } |\alpha| \geq |\beta|\,, \tag{8.3.8}$$

$$\widehat{\kappa}_L(\alpha,\beta\,;e_k) \;\leq\; \rho k^{3/2} \inf_v \widehat{\kappa}_L(\alpha,\beta\,;v) \;\; \textit{if } A_k \textit{ is nonsingular and } |\alpha| \leq |\beta|\,. \tag{8.3.9}$$

Proof. Note first that the nonsingularity conditions on A_0 and A_k ensure that 0 and ∞, respectively, are not eigenvalues of P, and hence that $v = e_1$ and $v = e_k$, respectively, yield linearizations.

Since $\widehat{\kappa}_L(\alpha,\beta\,;v)$ is invariant under scaling of v, we can set $\|v\|_2 = 1$. In view of the bounds in Lemma 8.3.1, the v-dependent term $\sqrt{|\alpha|^2\omega_X^2 + |\beta|^2\omega_Y^2}$ in the numerator of (8.2.9) is bounded below by $\min(\|A_0\|_2, \|A_k\|_2)\sqrt{|\alpha|^2 + |\beta|^2}$ for any such v, and bounded above by $k \max_i \|A_i\|_2 \sqrt{|\alpha|^2 + |\beta|^2}$ when $v = e_j$ for some j. Hence to prove (8.3.8) it suffices to show that

$$\max_{\|v\|_2=1} |\mathsf{p}(\alpha,\beta\,;v)| \leq \sqrt{k}\,|\mathsf{p}(\alpha,\beta\,;e_1)| \quad \text{for } |\alpha| \geq |\beta|. \tag{8.3.10}$$

This inequality is trivial for $\beta = 0$, so we can assume $\beta \neq 0$ and divide through by β^{k-1} to rewrite the desired inequality as

$$\max_{\|v\|_2=1} |\mathsf{p}(\lambda\,;v)| \leq \sqrt{k}\,|\mathsf{p}(\lambda\,;e_1)| \quad \text{for } |\lambda| \geq 1.$$

But this inequality follows from

$$|\mathsf{p}(\lambda\,; v)| = |\Lambda^T v| \le \|\Lambda\|_2 \le \sqrt{k}\,|\lambda^{k-1}| = \sqrt{k}\,|\mathsf{p}(\lambda\,; e_1)|.$$

The proof of (8.3.9) is entirely analogous. ∎

Theorem 8.3.2 says that for matrix polynomials with coefficient matrices of roughly equal norm (so ρ is of order 1), one of the two pencils with $v = e_1$ and $v = e_k$ will always give a *near optimal* condition number $\widehat{\kappa}_L$ for a given eigenvalue, in the sense that $\widehat{\kappa}_L$ differs from the minimal value by a factor of at most $\rho k^{3/2}$. Moreover, which pencil is nearly optimal depends only on whether the given eigenvalue is greater than or less than 1 in modulus. Note, however, that taking the wrong choice of $v = e_1$ or $v = e_k$ can be disastrous:

$$\widehat{\kappa}_L(0, \beta\,; e_1) = \infty\,, \qquad \widehat{\kappa}_L(\alpha, 0; e_k) = \infty \tag{8.3.11}$$

(and in these situations the pencils are not even linearizations); see the final example in Section 8.7.

For the quadratic polynomial $Q(\lambda) = \lambda^2 A + \lambda B + C$, the pencils corresponding to $v = e_1$ and $v = e_k (= e_2)$ are, respectively (from Table 3.3.1),

$$L_1(\lambda) = \lambda \begin{bmatrix} A & 0 \\ 0 & -C \end{bmatrix} + \begin{bmatrix} B & C \\ C & 0 \end{bmatrix}, \quad L_2(\lambda) = \lambda \begin{bmatrix} 0 & A \\ A & B \end{bmatrix} + \begin{bmatrix} -A & 0 \\ 0 & C \end{bmatrix}. \tag{8.3.12}$$

These pencils were analyzed by Tisseur [80], along with a companion form linearization (which belongs to $\mathbb{L}_1(Q)$ but not $\mathbb{DL}(Q)$). She showed that if $\|A\|_2 = \|B\|_2 = \|C\|_2 = 1$ then $\kappa_{L_1}(\lambda) \le \kappa_{L_2}(\lambda)$ for $|\lambda| \ge \sqrt{2}$ and $\kappa_{L_1}(\lambda) \ge \kappa_{L_2}(\lambda)$ for $|\lambda| \le 2^{-1/2}$. The analysis in Theorem 8.3.2 implies that analogous inequalities hold for arbitrary degrees k and arbitrary ρ. In fact, working directly from Lemma 8.3.1 we can show that

$$\begin{aligned} \widehat{\kappa}_L(\alpha, \beta\,; e_1) &\le \widehat{\kappa}_L(\alpha, \beta\,; e_k) \quad \text{if } |\alpha| \ge (\rho k)^{\frac{1}{k-1}} |\beta|, \\ \widehat{\kappa}_L(\alpha, \beta\,; e_k) &\le \widehat{\kappa}_L(\alpha, \beta\,; e_1) \quad \text{if } |\beta| \ge (\rho k)^{\frac{1}{k-1}} |\alpha|, \end{aligned}$$

with entirely analogous inequalities holding for $\kappa_L(\lambda)$.

Now we compare the optimal $\widehat{\kappa}_L(\alpha, \beta\,; v)$ with $\widehat{\kappa}_P(\alpha, \beta)$, the condition number of the eigenvalue for the original polynomial.

Theorem 8.3.3. *Let (α, β) be a simple eigenvalue of P. Then*

$$\frac{1}{\rho} \le \frac{\inf_v \widehat{\kappa}_L(\alpha, \beta\,; v)}{\widehat{\kappa}_P(\alpha, \beta)} \le k^2 \rho,$$

where the weights are chosen as $\omega_i \equiv \|A_i\|_2$ for $\widehat{\kappa}_P$ and as in (8.3.1) *for $\widehat{\kappa}_L$, and ρ is defined in* (8.3.7).

Proof. From Theorem 8.1.3,

$$\widehat{\kappa}_P(\alpha, \beta) = \frac{\left(\sum_{i=0}^k |\alpha|^{2i} |\beta|^{2(k-i)} \|A_i\|_2^2\right)^{1/2} \|y\|_2 \, \|x\|_2}{\left| y^* (\bar{\beta} \mathcal{D}_\alpha P - \bar{\alpha} \mathcal{D}_\beta P)|_{(\alpha,\beta)} x \right|}.$$

On the other hand, for $v = v_*$ in (8.3.6) we have, from Theorem 8.2.1,

$$\widehat{\kappa}_L(\alpha,\beta\,;v_*) = \frac{\sqrt{|\alpha|^2\|X\|_2^2 + |\beta|^2\|Y\|_2^2}\;\|\varLambda_{\alpha,\beta}\|_2\,\|y\|_2\,\|x\|_2}{\big|y^*(\bar\beta\mathcal{D}_\alpha P - \bar\alpha\mathcal{D}_\beta P)|_{(\alpha,\beta)}x\big|}\,. \tag{8.3.13}$$

If L is not a linearization for $v = v_*$ then we need to interpret v_* as an arbitrarily small perturbation of v_* for which L is a linearization. Using (8.3.2) and (8.3.3) and $\sum_{i=0}^k |\alpha|^{2i}|\beta|^{2(k-i)}\,\|A_i\|_2^2 \ge (|\alpha|^{2k}+|\beta|^{2k})\min(\|A_0\|_2,\|A_k\|_2)^2$, it is easy to see that

$$\frac{\widehat{\kappa}_L(\alpha,\beta\,;v_*)}{\widehat{\kappa}_P(\alpha,\beta)} \;\le\; \rho k^{3/2} f(\alpha,\beta),$$

where

$$f(\alpha,\beta) = \frac{\sqrt{|\alpha|^2+|\beta|^2}\,\Big(\sum_{i=1}^k |\alpha|^{2(i-1)}|\beta|^{2(k-i)}\Big)^{1/2}}{\sqrt{|\alpha|^{2k}+|\beta|^{2k}}}.$$

From (8.8.1) in Proposition 8.8.1 we have $f(\alpha,\beta) \le \sqrt{k}$. The upper bound follows since $\inf_v \widehat{\kappa}_L(\alpha,\beta\,;v) \le \widehat{\kappa}_L(\alpha,\beta\,;v_*)$. For the lower bound we have, for any v with $\|v\|_2 = 1$,

$$\begin{aligned}
\frac{\widehat{\kappa}_L(\alpha,\beta\,;v)}{\widehat{\kappa}_P(\alpha,\beta)} &= \frac{\sqrt{|\alpha|^2\|X\|_2^2+|\beta|^2\|Y\|_2^2}\;\|\varLambda_{\alpha,\beta}\|_2^2}{\Big(\sum_{i=0}^k |\alpha|^{2i}|\beta|^{2(k-i)}\|A_i\|_2^2\Big)^{1/2}\,|\mathsf{p}(\alpha,\beta\,;v)|}\\
&\ge \frac{\sqrt{|\alpha|^2+|\beta|^2}\,\min(\|A_0\|_2,\|A_k\|_2)\|\varLambda_{\alpha,\beta}\|_2}{\Big(\sum_{i=0}^k |\alpha|^{2i}|\beta|^{2(k-i)}\Big)^{1/2}\,\max_i\|A_i\|_2}\\
&\ge \frac{1}{\rho}\,\frac{\sqrt{|\alpha|^2+|\beta|^2}\,\Big(\sum_{i=1}^k |\alpha|^{2(i-1)}|\beta|^{2(k-i)}\Big)^{1/2}}{\Big(\sum_{i=0}^k |\alpha|^{2i}|\beta|^{2(k-i)}\Big)^{1/2}} =: \frac{1}{\rho}g(\alpha,\beta),
\end{aligned}$$

since $|\mathsf{p}(\alpha,\beta\,;v)| \le \|\varLambda_{\alpha,\beta}\|_2$ by the Cauchy–Schwarz inequality. From (8.8.2), $g(\alpha,\beta) \ge 1$, and the lower bound follows. ∎

Finally we state the analogs of Theorem 8.3.2 and 8.3.3 for $\kappa_L(\lambda)$. Keep in mind that ρ is the quantity defined in (8.3.7).

Theorem 8.3.4. *Let λ be a simple, finite, nonzero eigenvalue of P and consider pencils $L \in \mathbb{DL}(P)$. Take the weights* (8.3.1) *for κ_L. Then*

$$\kappa_L(\lambda\,;e_1) \;\le\; \rho k^{3/2}\inf_v \kappa_L(\lambda\,;v) \quad \textit{if } A_0 \textit{ is nonsingular and } |\lambda| \ge 1\,, \tag{8.3.14}$$

$$\kappa_L(\lambda\,;e_k) \;\le\; \rho k^{3/2}\inf_v \kappa_L(\lambda\,;v) \quad \textit{if } A_k \textit{ is nonsingular and } |\lambda| \le 1\,. \tag{8.3.15}$$

Proof. The proof is entirely analogous to that of Theorem 8.3.2. ∎

Theorem 8.3.5. *Let λ be a simple, finite, nonzero eigenvalue of P. Then*

$$\left(\frac{2\sqrt{k}}{k+1}\right)\frac{1}{\rho} \;\le\; \frac{\inf_v \kappa_L(\lambda\,;v)}{\kappa_P(\lambda)} \;\le\; k^2\rho\,,$$

where the weights are chosen as $\omega_i \equiv \|A_i\|_2$ for κ_P and as in (8.3.1) *for L.*

Proof. The proof is very similar to that of Theorem 8.3.3, but with slightly different f and g having the form of f_3 and f_4 in Proposition 8.8.1. $\square$

Theorems 8.3.3 and 8.3.5 show that for polynomials whose coefficient matrices do not vary too much in norm, the best conditioned linearization in $\mathbb{DL}(P)$ for a particular eigenvalue is about as well conditioned as P itself for that eigenvalue, to within a small constant factor. This is a rather surprising result, given that the condition numbers $\widehat{\kappa}_L(\alpha,\beta)$ and $\kappa_L(\lambda)$ permit *arbitrary* perturbations in the $\mathbb{DL}(P)$-pencils $L(\lambda) = \lambda X + Y$ that do not respect the zero and repeated block structure of X and Y (for two particular instances of this block structure with $k = 2$ see (8.3.12)). Under the same assumptions on the $\|A_i\|_2$, by combining Theorems 8.3.2 and 8.3.3 or Theorems 8.3.4 and 8.3.5 we can conclude that, for any given eigenvalue, one of the two pencils with $v = e_1$ and $v = e_k$ will be about as well conditioned as P itself for that eigenvalue.

8.3.1 Several Eigenvalues

Suppose now that several eigenvalues $(\alpha_1, \beta_1), \ldots, (\alpha_r, \beta_r)$ are of interest and that neither $|\alpha_i| \geq |\beta_i|$ for all i nor $|\alpha_i| \leq |\beta_i|$ for all i. A reasonable way to define a single pencil that is best for all these eigenvalues is by maximizing the 2-norm of the r-vector of the reciprocals of the eigenvalue condition numbers for the pencil. This vector can be written, using Theorem 8.2.1, as

$$\begin{aligned} &\operatorname{diag}\big((|\alpha_i|^2\omega_X^2 + |\beta_i|^2\omega_Y^2)^{1/2}\,\|\Lambda_{\alpha_i,\beta_i}\|_2^2\|y_i\|_2\|x_i\|_2\big)^{-1} \\ &\times \operatorname{diag}(|y_i^*(\bar\beta_i\mathcal{D}_\alpha P - \bar\alpha_i\mathcal{D}_\beta P)|_{(\alpha_i,\beta_i)}x_i|) \begin{bmatrix} \Lambda_{\alpha_1,\beta_1}^T \\ \vdots \\ \Lambda_{\alpha_r,\beta_r}^T \end{bmatrix} v =: Bv. \end{aligned}$$

Assume that $\rho = O(1)$, so that ω_X and ω_Y in (8.3.1) are roughly constant in $\|v\|_2$. Then we can set $\omega_X = \omega_Y = 1$ and define the optimal v as the right singular vector corresponding to the largest singular value of B. This approach requires knowledge of the eigenvectors x_i and y_i as well as the λ_i. If the eigenvectors are not known then we can simplify B further to

$$\operatorname{diag}\Big((|\alpha_i|^2 + |\beta_i|^2)^{1/2}\,\|\Lambda_{\alpha_i,\beta_i}\|_2^2\Big)^{-1} \begin{bmatrix} \Lambda_{\alpha_1,\beta_1}^T \\ \vdots \\ \Lambda_{\alpha_r,\beta_r}^T \end{bmatrix}.$$

So far we have implicitly assumed that we have a good estimate of the eigenvalues of interest. Suppose, instead, that we know only a region S of the complex plane in which the eigenvalues of interest lie. In this case a natural approach is to try to minimize the v-dependent part of the eigenvalue condition number over S. Continuing to assume $\rho = O(1)$, and working now with $\kappa_L(\lambda; v)$, the problem becomes to find the v that achieves the maximum in the problem

$$\max_{\|v\|_2=1} \min_{\lambda\in S} |\mathsf{p}(\lambda; v)|.$$

This uniform (or Chebyshev) complex approximation problem can be expressed as a semi-infinite programming problem and solved by numerical methods for such problems [70, Sec. 2.3].

8.4 Quadratic Polynomials

We now concentrate our attention on quadratic polynomials $Q(\lambda) = \lambda^2 A + \lambda B + C$, since these are in practice the most important polynomials of degree 2 or higher. For brevity write

$$a = \|A\|_2, \quad b = \|B\|_2, \quad c = \|C\|_2. \tag{8.4.1}$$

The quantity ρ in Theorems 8.3.2–8.3.5 is now

$$\rho = \frac{\max(a,b,c)}{\min(a,c)}.$$

Clearly, ρ is of order 1 if

$$b \lesssim \max(a,c) \quad \text{and} \quad a \approx c.$$

If these conditions are not satisfied then we can consider scaling Q to try to improve ρ. Write $\lambda = \mu\gamma$, $\gamma \in \mathbb{R}$ and

$$Q(\lambda) = \lambda^2 A + \lambda B + C = \mu^2(\gamma^2 A) + \mu(\gamma B) + C =: \mu^2 \widetilde{A} + \mu\widetilde{B} + \widetilde{C} =: \widetilde{Q}(\mu)\,. \tag{8.4.2}$$

The γ that minimizes $\max(\|\widetilde{A}\|_2/\|\widetilde{B}\|_2, \|\widetilde{C}\|_2/\|\widetilde{B}\|_2) = \max(\gamma a/b, c/(\gamma b))$ is easily seen to be

$$\gamma = \sqrt{c/a}, \tag{8.4.3}$$

and it yields

$$\|\widetilde{A}\|_2 = c, \quad \|\widetilde{B}\|_2 = b\sqrt{c/a}, \quad \|\widetilde{C}\|_2 = c.$$

Hence, for the scaled problem,

$$\rho = \max(1, b/\sqrt{ac}).$$

This scaling is intended to improve the conditioning of the linearizations, but what does it do to the conditioning of the quadratic itself? It is easy to see that $\kappa_P(\lambda)$ is invariant under this scaling when $\omega_i = \|A_i\|_2$, but that $\widehat{\kappa}_P(\alpha,\beta)$ is scale-dependent. We note that the scaling (8.4.2) and (8.4.3) is used by Fan, Lin, and Van Dooren [23]; see Section 8.6.

With these observations Theorems 8.3.4 and 8.3.5 can be combined and specialized as follows.

Theorem 8.4.1. *Let λ denote a simple eigenvalue of $Q(\lambda) = \lambda^2 A + \lambda B + C$ or of the scaled quadratic $\widetilde{Q}$ defined by* (8.4.2) *and* (8.4.3)*. Take the weights* (8.3.1) *for $\kappa_L(\lambda)$. With the notation* (8.4.1)*, assume that either*

- *$b \lesssim \max(a,c)$ and $a \approx c$, in which case let $P = Q$ and $L \in \mathbb{DL}(Q)$, or*
- *$b \lesssim \sqrt{ac}$, in which case let $P = \widetilde{Q}$ and $L \in \mathbb{DL}(\widetilde{Q})$.*

Then if C is nonsingular and $|\lambda| \ge 1$, the linearization with $v = e_1$ has $\kappa_L(\lambda; e_1) \approx \kappa_P(\lambda)$, while if A is nonsingular and $|\lambda| \le 1$, the linearization with $v = e_2$ has $\kappa_L(\lambda; e_2) \approx \kappa_P(\lambda)$.

If we think of Q as representing a mechanical system with damping, then the near-optimality of the $v = e_1$ and $v = e_2$ linearizations holds for Q that are not too heavily damped. One class of Q for which $b \lesssim \sqrt{ac}$ automatically holds is the elliptic Q [36], [48]: those for which A is Hermitian positive definite, B and C are Hermitian, and $(x^*Bx)^2 < 4(x^*Ax)(x^*Cx)$ for all nonzero $x \in \mathbb{C}^n$.

An analog of Theorem 8.4.1 for $\widehat{\kappa}_L(\alpha, \beta)$ can be obtained from Theorems 8.3.2 and 8.3.3.

8.5 Linearizing the Reversal of P

Consider the quadratic $Q(\lambda) = \lambda^2 A + \lambda B + C$ and its "reversal" $\operatorname{rev} Q(\lambda) = \lambda^2 C + \lambda B + A$. Since the eigenvalues of $\operatorname{rev} Q$ are just the reciprocals of those of Q, it is natural to wonder whether this relationship can be exploited to improve the conditioning of an eigenvalue computation. Tisseur [80, Lem. 10] shows that if λ is a simple, finite, nonzero eigenvalue of Q and $\mu = 1/\lambda$ the corresponding simple eigenvalue of $\operatorname{rev} Q$ then, with the weights (8.3.1), $\kappa_{\widetilde{L}_1}(\mu) = \kappa_{L_2}(\lambda)$ and $\kappa_{\widetilde{L}_2}(\mu) = \kappa_{L_1}(\lambda)$, where L_1 and L_2 are the pencils corresponding to $v = e_1$ and $v = e_2$ given in (8.3.12) and $\widetilde{L}_1$ and $\widetilde{L}_2$ are the corresponding pencils for $\operatorname{rev} Q$. In essence this result says that one cannot improve the condition of an eigenvalue of a linearization by regarding it as the reciprocal of an eigenvalue of the reversed quadratic. In this section we generalize this result in three respects: to any vector v (not just $v = e_1$ or e_2), to arbitrary degree polynomials, and to zero and infinite eigenvalues.

For P of degree k, define

$$\operatorname{rev} P(\lambda) = \lambda^k P(1/\lambda)\,;$$

as its name suggests, $\operatorname{rev} P$ is just the polynomial obtained by reversing the order of the coefficient matrices of P. Let $L(\lambda) = \lambda X + Y$ be the unique pencil in $\mathbb{DL}(P)$ with ansatz vector v and $\widetilde{L}(\lambda) = \lambda \widetilde{X} + \widetilde{Y}$ the unique pencil in $\mathbb{DL}(\operatorname{rev} P)$ with ansatz vector Rv, where

$$R = \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix} \in \mathbb{R}^{k \times k}.$$

Lemma 8.5.1. *L is a linearization for P if and only if $\widetilde{L}$ is a linearization for* $\operatorname{rev} P$.

Proof. The roots of the v-polynomial $\mathsf{p}(x\,; Rv)$ are the reciprocals of the roots of $\mathsf{p}(x\,; v)$, while the eigenvalues of $\operatorname{rev} P$ are the reciprocals of the eigenvalues of P. The result now follows from the Eigenvalue Exclusion Theorem 4.3.1. ∎

We now work with the condition number $\widehat{\kappa}_L(\alpha, \beta)$, since reciprocal pairs (including 0 and ∞) are so conveniently represented in homogeneous form by (α, β) and (β, α). First observe that (α, β) is an eigenvalue of P with right and left eigenvectors x and y if and only if (β, α) is an eigenvalue of $\operatorname{rev} P$ with right and left eigenvectors x and y. Also note that in homogeneous variables $\operatorname{rev} P(\alpha, \beta) = P(\beta, \alpha)$.

Lemma 8.5.2. *If the weights ω_X and ω_Y for L and weights $\omega_{\widetilde{X}}$ and $\omega_{\widetilde{Y}}$ for $\widetilde{L}$ satisfy the "crossover" equalities $\omega_X = \omega_{\widetilde{Y}}$ and $\omega_Y = \omega_{\widetilde{X}}$, then $\widehat{\kappa}_L(\alpha, \beta) = \widehat{\kappa}_{\widetilde{L}}(\beta, \alpha)$.*

Proof. We have, from (8.2.9),

$$\begin{aligned}
\widehat{\kappa}_L(\alpha,\beta) &= \frac{\sqrt{|\alpha|^2\omega_X^2+|\beta|^2\omega_Y^2}}{|\mathsf{p}(\alpha,\beta\,;v)|}\cdot\frac{\|\Lambda_{\alpha,\beta}\|_2^2\,\|y\|_2\,\|x\|_2}{\left|y^*(\bar{\beta}\mathcal{D}_\alpha P-\bar{\alpha}\mathcal{D}_\beta P)|_{(\alpha,\beta)}x\right|}\,,\\
\kappa_{\widetilde{L}}(\beta,\alpha) &= \frac{\sqrt{|\beta|^2\omega_{\widetilde{X}}^2+|\alpha|^2\omega_{\widetilde{Y}}^2}}{|\mathsf{p}(\beta,\alpha;Rv)|}\cdot\frac{\|\Lambda_{\beta,\alpha}\|_2^2\,\|y\|_2\,\|x\|_2}{\left|y^*(\bar{\alpha}\mathcal{D}_\alpha\mathrm{rev}\,P-\bar{\beta}\mathcal{D}_\beta\mathrm{rev}\,P)|_{(\beta,\alpha)}x\right|}\,.
\end{aligned}$$

We show that each of the four terms in the first expression equals the corresponding term in the second expression. The assumptions on the weights clearly imply equality of the square root terms. Next, $\Lambda_{\beta,\alpha}=R\Lambda_{\alpha,\beta}$, so $\Lambda_{\beta,\alpha}$ and $\Lambda_{\alpha,\beta}$ have the same 2-norm, while $\mathsf{p}(\alpha,\beta\,;v)\equiv\mathsf{p}(\beta,\alpha;Rv)$. Finally,

$$\begin{aligned}
(\bar{\alpha}\mathcal{D}_\alpha\mathrm{rev}\,P-\bar{\beta}\mathcal{D}_\beta\mathrm{rev}\,P)|_{(\beta,\alpha)} &= \bar{\alpha}(\mathcal{D}_\alpha\mathrm{rev}\,P)|_{(\beta,\alpha)}-\bar{\beta}(\mathcal{D}_\beta\mathrm{rev}\,P)|_{(\beta,\alpha)}\\
&= \bar{\alpha}(\mathcal{D}_\beta P)|_{(\alpha,\beta)}-\bar{\beta}(\mathcal{D}_\alpha P)|_{(\alpha,\beta)}\\
&= -(\bar{\beta}\mathcal{D}_\alpha P-\bar{\alpha}\mathcal{D}_\beta P)|_{(\alpha,\beta)},
\end{aligned}$$

which implies the equality of the final two denominator terms. ∎

Do the crossover conditions $\omega_X=\omega_{\widetilde{Y}}$ and $\omega_Y=\omega_{\widetilde{X}}$ hold for the natural choice of weights $\omega_X\equiv\|X\|_2$, $\omega_Y\equiv\|Y\|_2$? The next lemma shows that they do, by establishing an even stronger relationship between L and $\widetilde{L}$.

Lemma 8.5.3. *We have*

$$\widetilde{L}(\lambda)=(R\otimes I_n)\mathrm{rev}\,L(\lambda)(R\otimes I_n), \tag{8.5.1}$$

and so $\widetilde{X}=(R\otimes I_n)Y(R\otimes I_n)$ *and* $\widetilde{Y}=(R\otimes I_n)X(R\otimes I_n)$. *Hence* $\|\widetilde{X}\|=\|Y\|$ *and* $\|\widetilde{Y}\|=\|X\|$ *for any unitarily invariant norm.*

Proof. $\widetilde{L}$ is *defined* as the unique pencil in $\mathbb{DL}(\mathrm{rev}\,P)=\mathbb{L}_1(\mathrm{rev}\,P)\cap\mathbb{L}_2(\mathrm{rev}\,P)$ corresponding to the ansatz vector Rv. Therefore to establish (8.5.1) it suffices to show that the pencil $(R\otimes I_n)\mathrm{rev}\,L(\lambda)(R\otimes I_n)$ belongs to both $\mathbb{L}_1(\mathrm{rev}\,P)$ and $\mathbb{L}_2(\mathrm{rev}\,P)$ with right/left ansatz vector Rv. The other results then follow.

Recall that $\mathrm{rev}\,P(\lambda)=\lambda^kP(1/\lambda)$ and note that $\lambda^{k-1}\Lambda(1/\lambda)=R\Lambda$, where $\Lambda(r)$ is defined in (1.1.5). If $L\in\mathbb{L}_1(P)$ with right ansatz vector v then

$$\begin{aligned}
&L(\lambda)\cdot(\Lambda\otimes I_n)=v\otimes P(\lambda)\\
&\qquad\Rightarrow\quad L(1/\lambda)\cdot(\Lambda(1/\lambda)\otimes I_n)=v\otimes P(1/\lambda)\\
&\qquad\Rightarrow\quad \lambda L(1/\lambda)\cdot(\lambda^{k-1}\Lambda(1/\lambda)\otimes I_n)=v\otimes\lambda^kP(1/\lambda)\\
&\qquad\Rightarrow\quad \mathrm{rev}\,L(\lambda)\cdot(R\Lambda\otimes I_n)=v\otimes\mathrm{rev}\,P(\lambda)\\
&\qquad\Rightarrow\quad (R\otimes I_n)\mathrm{rev}\,L(\lambda)(R\otimes I_n)\cdot(\Lambda\otimes I_n)=(R\otimes I_n)(v\otimes\mathrm{rev}\,P(\lambda))\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\;\;=Rv\otimes\mathrm{rev}\,P(\lambda)\,,
\end{aligned}$$

which means that the pencil $(R\otimes I_n)\mathrm{rev}\,L(\lambda)(R\otimes I_n)$ is in $\mathbb{L}_1(\mathrm{rev}\,P)$ with right ansatz vector Rv.

In a similar manner it can be shown that $L\in\mathbb{L}_2(P)$ with left ansatz vector v implies that the pencil $(R\otimes I_n)\mathrm{rev}\,L(\lambda)(R\otimes I_n)$ is in $\mathbb{L}_2(\mathrm{rev}\,P)$ with left ansatz vector Rv. ∎

Combining the previous three lemmas we obtain the following generalization of Tisseur [80, Lem. 10].

Theorem 8.5.4. *Let (α, β) be a simple eigenvalue of P, so that (β, α) is a simple eigenvalue of* $\mathrm{rev}P$. *Suppose $L \in \mathbb{DL}(P)$ with ansatz vector v is a linearization of P. Then $\widetilde{L} \in \mathbb{DL}(\mathrm{rev}P)$ with ansatz vector Rv is a linearization of* $\mathrm{rev}P$ *and, if the weights are chosen as in* (8.3.1), $\widehat{\kappa}_L(\alpha, \beta) = \widehat{\kappa}_{\widetilde{L}}(\beta, \alpha)$.

An analog of Theorem 8.5.4 stating that $\kappa_L(\lambda) = \kappa_{\widetilde{L}}(1/\lambda)$ for finite, nonzero λ can also be derived by similar arguments.

8.6 Companion Linearizations

Recall the two companion form linearizations $C_1(\lambda) = \lambda X_1 + Y_1$ and $C_2(\lambda) = \lambda X_2 + Y_2$ from Definition 1.1.5, where

$$X_1 = X_2 = \mathrm{diag}(A_k, I_n, \dots, I_n),$$

$$Y_1 = \begin{bmatrix} A_{k-1} & A_{k-2} & \dots & A_0 \\ -I_n & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -I_n & 0 \end{bmatrix}, \quad \text{and} \quad Y_2 = \begin{bmatrix} A_{k-1} & -I_n & \dots & 0 \\ A_{k-2} & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & -I_n \\ A_0 & 0 & \dots & 0 \end{bmatrix}.$$

In Chapter 2 these pencils were the two key examples motivating the definition of the pencil spaces $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ that have been the central players throughout this thesis. In addition they have historically been the linearizations most used in practice. Thus it is entirely appropriate to give a special analysis of the conditioning of these two particular linearizations, and to compare their behavior with that of P and of suitable linearizations from $\mathbb{DL}(P)$. Recall that $C_1 \in \mathbb{L}_1(P)$ with right ansatz vector $v = e_1$, $C_2 \in \mathbb{L}_2(P)$ with left ansatz vector $w = e_1$, but neither pencil is in $\mathbb{DL}(P)$.

Our first result shows that it suffices to analyze the conditioning of C_1, because any results about the conditioning of C_1 translate to C_2 simply by transposing the coefficient matrices A_i.

Lemma 8.6.1. *Let λ, or (α, β) in homogeneous form, be a simple eigenvalue of P, and take $\omega_i = \|A_i\|_2$. Then*

$$\widehat{\kappa}_P(\alpha, \beta) = \widehat{\kappa}_{P^T}(\alpha, \beta), \qquad \kappa_P(\lambda) = \kappa_{P^T}(\lambda).$$

Moreover,

$$\widehat{\kappa}_{C_2(P)}(\alpha, \beta) = \widehat{\kappa}_{C_1(P^T)}(\alpha, \beta), \qquad \kappa_{C_2(P)}(\lambda) = \kappa_{C_1(P^T)}(\lambda),$$

where $C_i(P)$, $i = 1, 2$, denotes the ith companion linearization for P, and P^T denotes the polynomial obtained by transposing each coefficient matrix A_i.

Proof. If (λ, x, y) is an eigentriple for P then $(\lambda, \overline{y}, \overline{x})$ is an eigentriple for P^T. The first two equalities follow by considering the formulae (8.1.2) and (8.1.5). It is easy to see that $C_2(P) = C_1(P^T)^T$. The second pair of equalities are therefore special cases of the first. ∎

For the rest of the section we work with λ and $\kappa(\lambda)$; for (α,β) and $\widehat{\kappa}(\alpha,\beta)$ analogous results hold. We first obtain a formula for left eigenvectors w^* of C_1.

Lemma 8.6.2. *The vector $y \in \mathbb{C}^n$ is a left eigenvector of P corresponding to a simple, finite, nonzero eigenvalue λ if and only if*

$$w = \begin{bmatrix} I \\ (\lambda A_k + A_{k-1})^* \\ \vdots \\ (\lambda^{k-1}A_k + \lambda^{k-2}A_{k-1} + \cdots + A_1)^* \end{bmatrix} y \tag{8.6.1}$$

is a left eigenvector of C_1 corresponding to λ.

Proof. Since C_1 is a linearization of P, λ is a simple eigenvalue of C_1. The proof therefore consists of a direct verification that $w^*C_1(\lambda) = 0$. ∎

Lemma 8.6.2 shows that, even though $C_1 \notin \mathbb{L}_2(P)$, a left eigenvector of P can be recovered from one of C_1 — simply by reading off the *leading* n components.

Since $C_1 \in \mathbb{L}_1(P)$, we know that the right eigenvectors z and x of C_1 and P are related by $z = \Lambda \otimes x$. Evaluating (8.2.11) (which holds for any member of $\mathbb{L}_1(P)$) with $L = C_1$ at an eigenvalue λ, then multiplying on the left by w^* and on the right by $1 \otimes x = x$, we obtain

$$w^*C_1'(\lambda)z \;=\; w^*(v \otimes P'(\lambda)x).$$

Using the formula (8.6.1) for w and the fact that $v = e_1$ gives

$$w^*C_1'(\lambda)z = y^*P'(\lambda)x.$$

By applying Theorem 8.1.1 to C_1 we obtain the following analog of Theorem 8.2.2.

Theorem 8.6.3. *Let λ be a simple, finite, nonzero eigenvalue of P with right and left eigenvectors x and y, respectively. Then, for the first companion linearization $C_1(\lambda) = \lambda X_1 + Y_1$,*

$$\kappa_{C_1}(\lambda) = \frac{(|\lambda|\omega_{X_1} + \omega_{Y_1})\,\|w\|_2\,\|\Lambda\|_2\,\|x\|_2}{|\lambda|\,|y^*P'(\lambda)x|},$$

where w is given by (8.6.1).

Now we can compare the condition number of the first companion form with that of P. We have

$$\frac{\kappa_{C_1}(\lambda)}{\kappa_P(\lambda)} \;=\; \frac{\|w\|_2}{\|y\|_2}\cdot\frac{(|\lambda|\omega_{X_1} + \omega_{Y_1})\,\|\Lambda\|_2}{\sum_{i=0}^{k}|\lambda|^i\omega_i}.$$

We choose the weights $\omega_{X_1} = \|X_1\|_2$, $\omega_{Y_1} = \|Y_1\|_2$, and $\omega_i = \|A_i\|_2$ in (8.1.2), and consequently need bounds on the norms of X_1 and Y_1. These are provided by the next lemma, which is similar to Lemma 8.3.1.

Lemma 8.6.4. *For $C_1(\lambda) = \lambda X_1 + Y_1$ we have $\|X_1\|_2 = \max(\|A_k\|_2, 1)$ and*

$$\max\Bigl(1, \max_{i=0:\,k-1}\|A_i\|_2\Bigr) \le \|Y_1\|_2 \le k\max\Bigl(1, \max_{i=0:\,k-1}\|A_i\|_2\Bigr). \tag{8.6.2}$$

Proof. Straightforward, using (8.3.5). ∎

For notational simplicity we will now concentrate on the quadratic case, $k = 2$. With the notation (8.4.1), we have

$$\frac{\psi}{2^{1/2}} \frac{\|w\|_2}{\|y\|_2} \leq \frac{\kappa_{C_1}(\lambda)}{\kappa_P(\lambda)} \leq 2\psi \frac{\|w\|_2}{\|y\|_2} \tag{8.6.3}$$

where

$$\psi = \frac{(1+|\lambda|)\big(\max(a,1)|\lambda| + \max(b,c,1)\big)}{a|\lambda|^2 + b|\lambda| + c} \geq 1$$

and

$$\frac{\|w\|_2}{\|y\|_2} = \frac{\left\| \begin{bmatrix} I \\ (\lambda A + B)^* \end{bmatrix} y \right\|_2}{\|y\|_2} = \frac{\left\| \begin{bmatrix} I \\ (\lambda^{-1} C)^* \end{bmatrix} y \right\|_2}{\|y\|_2} \tag{8.6.4}$$

satisfies

$$1 \leq \frac{\|w\|_2}{\|y\|_2} \leq \min\big((1 + (|\lambda|a + b)^2)^{1/2}, (1 + c^2/|\lambda|^2)^{1/2}\big).$$

Therefore $\kappa_{C_1}(\lambda)$ will be of the same order of magnitude as $\kappa_P(\lambda)$ only if both ψ and $\|w\|_2/\|y\|_2$ are of order 1. It is difficult to characterize when these conditions hold. However, it is clear that, unlike for the $\mathbb{DL}(P)$ linearizations, the conditioning of C_1 is affected by scaling $A_i \leftarrow \gamma A_i$, $i = 0\colon k$, as might be expected in view of the mixture of identity matrices and A_i that make up the blocks of X_1 and Y_1. Indeed if $a, b, c \ll 1$ then $\psi \gg 1$, while if $a, b, c \gg |\lambda| \geq 1$ then $\|w\|_2/\|y\|_2 \gg 1$, unless y is nearly a null vector for $(\lambda A + B)^*$ and C^*. The only straightforward conditions that guarantee $\kappa_{C_1}(\lambda) \approx \kappa_P(\lambda)$ are $a \approx b \approx c \approx 1$: then $\psi \approx 1$ and one of the two expressions for $\|w\|_2/\|y\|_2$ in (8.6.4) is clearly of order 1 (the first if $|\lambda| \leq 1$, otherwise the second). The predilection of the first companion form for coefficient matrices of unit 2-norm was shown from a different viewpoint by Tisseur [80, Thm. 7]: she proves that when $a = b = c = 1$, applying a backward stable solver to the companion pencil is backward stable for the original quadratic.

It is natural to scale the problem to try to bring the 2-norms of A, B, and C close to 1. The scaling of Fan, Lin, and Van Dooren [23], which was motivated by backward error considerations, has precisely this aim. It converts $Q(\lambda) = \lambda^2 A + \lambda B + C$ to $\widetilde{Q}(\mu) = \mu^2 \widetilde{A} + \mu \widetilde{B} + \widetilde{C}$, where

$$\lambda = \gamma\mu, \quad Q(\lambda)\delta = \mu^2(\gamma^2 \delta A) + \mu(\gamma \delta B) + \delta C \equiv \widetilde{Q}(\mu), \tag{8.6.5a}$$

$$\gamma = \sqrt{c/a}, \quad \delta = 2/(c + b\gamma). \tag{8.6.5b}$$

This is the same scaling factor γ we used in Section 8.4, combined with the multiplication of each coefficient matrix by δ.

Now we compare $\kappa_{C_1}(\lambda)$ with $\kappa_L(\lambda\,; v_*)$, where v_* for λ is defined analogously to v_* for (α, β) in (8.3.6) by

$$v_* = \frac{\overline{\Lambda}}{\|\Lambda\|_2}, \qquad |\mathsf{p}(\lambda\,; v_*)| = \|\Lambda\|_2\,.$$

We have, from (8.2.12),

$$\kappa_L(\lambda\,; v_*) = \frac{(|\lambda|\omega_X + \omega_Y)\|A\|_2\,\|y\|_2\,\|x\|_2}{|\lambda|\;|y^*P'(\lambda)x|},$$

and so

$$\frac{\kappa_{C_1}(\lambda)}{\kappa_L(\lambda\,; v_*)} = \frac{\|w\|_2}{\|y\|_2}\cdot\frac{|\lambda|\omega_{X_1}+\omega_{Y_1}}{|\lambda|\omega_X+\omega_Y}.$$

Again, specializing to $k = 2$, and using Lemmas 8.3.1 and 8.6.4, we have

$$\begin{aligned}\frac{\|w\|_2}{\|y\|_2}\cdot\frac{\bigl(\max(a,1)|\lambda|+\max(b,c,1)\bigr)}{2^{3/2}\max(a,b,c)(1+|\lambda|)} &\le \frac{\kappa_{C_1}(\lambda)}{\kappa_L(\lambda\,; v_*)} \\ &\le \frac{\|w\|_2}{\|y\|_2}\cdot\frac{\bigl(\max(a,1)|\lambda|+2\max(b,c,1)\bigr)}{a|\lambda|+c}.\end{aligned} \tag{8.6.6}$$

If $a \approx b \approx c \approx 1$ then we can conclude that $\kappa_{C_1}(\lambda) \approx \kappa_L(\lambda\,; v_*)$. However, $\kappa_{C_1}(\lambda) \gg \kappa_L(\lambda\,; v_*)$ if $\|w\|_2/\|y\|_2 \gg 1$ or if (for example) $a, b, c \ll 1$.

Our results for the companion forms are not as neat as those in Section 8.3 for the $\mathbb{DL}(P)$-linearizations, which focus attention on a single, easily computed or estimated, scalar parameter ρ. The conditioning of the companion forms relative to P and to the class $\mathbb{DL}(P)$ depends on both (a) the ratios of norms of left eigenvectors of C_1 and P and (b) rational functions of the coefficient matrix norms and λ. It does not seem possible to bound the norm ratio in a useful way a priori. Therefore the only easily checkable conditions that we can identify under which the companion forms can be guaranteed to be optimally conditioned are $\|A_i\|_2 \approx 1$, $i = 0\colon k$ (our proof of this fact for $k = 2$ is easily seen to generalize to arbitrary k).

Finally, we note that the bounds (8.6.3) and (8.6.6) remain true when "λ" is replaced by "α, β", with just minor changes to the constants.

8.7 Numerical Experiments

In this final section we illustrate the theory developed in this chapter on four quadratic eigenvalue problems. Experiments were performed in MATLAB 7, for which the unit roundoff is $2^{-53} \approx 1.1 \times 10^{-16}$. To obtain the angular error $\theta\bigl((\alpha,\beta),(\widetilde{\alpha},\widetilde{\beta})\bigr)$ for a computed eigenvalue $(\widetilde{\alpha},\widetilde{\beta})$, we took as exact eigenvalue (α,β) the value computed in MATLAB's VPA arithmetic at 40 digit precision. In all the figures the x-axis is the eigenvalue index and the eigenvalues are sorted in increasing order of absolute value. We compare the $\widehat{\kappa}$ condition numbers of the quadratic Q, the first companion form, and the $\mathbb{DL}(Q)$ linearizations with $v = e_1$ and $v = e_2$. All four problems have real symmetric coefficient matrices, so we know from Lemma 8.6.1 that the second companion form has exactly the same condition numbers as the first companion form. In three of the problems we apply the scaling given by (8.6.5). Table 8.7.1 reports the problem sizes, the coefficient matrix norms, and the values of ρ in (8.3.7) before and after scaling.

Our first problem shows the benefits of scaling. It comes from applying the Galerkin method to a PDE describing the wave motion of a vibrating string with

Table 8.7.1: Problem statistics.

Problem	Wave		Nuclear		Mass-spring	Acoustics	
n	25		8		50	107	
	Unscaled	Scaled	Unscaled	Scaled	Unscaled	Unscaled	Scaled
$\|A\|_2$	1.57e0	1.85e0	2.35e8	1.18e0	1.00e0	1.00e0	2.00e0
$\|B\|_2$	3.16e0	1.49e-1	4.35e10	8.21e-1	3.20e2	5.74e-2	3.64e-5
$\|C\|_2$	9.82e2	1.85e0	1.66e13	1.18e0	5.00e0	9.95e6	2.00e0
ρ	6.25e2	1.00e0	7.06e4	1.00e0	3.20e2	9.95e6	1.00e0

clamped ends in a spatially inhomogeneous environment [26], [36]. The quadratic Q is elliptic; the eigenvalues are nonreal and have absolute values in the interval $[1, 25]$. Figure 8.7.1 shows the condition numbers $\widehat{\kappa}_L(\alpha,\beta)$ for the $\mathbb{DL}(Q)$ linearization with $v = e_1$ and the first companion linearization, the condition number $\widehat{\kappa}_P(\alpha,\beta)$ for Q, and the angular errors in the eigenvalues computed by applying the QZ algorithm to the two linearizations. Figure 8.7.2 shows the corresponding information for the scaled problem. Since the eigenvalues are all of modulus at least 1, we know from Theorem 8.3.3 that for every eigenvalue, the $\mathbb{DL}(Q)$ linearization with $v = e_1$ has condition number within a factor $4\rho = 2500$ of the condition number for Q. The actual ratios are between 3.5 and 266. Since this problem is elliptic, we know from Theorem 8.4.1 that for the scaled problem, whose eigenvalues lie between 0.04 and 1 in modulus, the $\mathbb{DL}(\widetilde{Q})$ linearization with $v = e_2$ will have condition number similar to that of $\widetilde{Q}$ for every eigenvalue. This is confirmed by Figure 8.7.2; the maximum ratio of condition numbers is 3.3. The benefit of the smaller condition numbers after scaling is clear from the figures: the angular error of the computed eigenvalues is smaller by a factor roughly equal to the reduction in condition number. The behaviour of the companion linearization is very similar to that of the $\mathbb{DL}(Q)$ linearizations, and this is predicted by our theory since the term $\psi\|w\|_2/\|y\|_2$ in (8.6.3) varies from 3.7 to 511 without scaling and only 1.0 to 4.5 with scaling.

The next problem is a simplified model of a nuclear power plant [43], [83]. There are 2 real and 14 nonreal eigenvalues, with absolute values in the interval $(17, 362)$. Since $\rho = 7 \times 10^4$, it is not surprising that the $\mathbb{DL}(Q)$ linearization with $v = e_1$ has eigenvalue condition numbers up to 369 times as large as those of Q, as Figure 8.7.3 indicates. Although the problem is not elliptic, $\|B\|_2 \le \sqrt{\|A\|_2\|C\|_2}$, and so our theory says that scaling will make the $\mathbb{DL}(Q)$ linearization with $v = e_2$ (since the scaled eigenvalues have modulus at most 1) optimally conditioned. This prediction is confirmed in Figure 8.7.4. Scaling also brings a dramatic improvement in the conditioning and accuracy of the companion linearization; again, this is predicted by our theory since the scaled problem has coefficient matrices of norm approximately 1, and the magnitude of the reduction is explained by the term $\psi\|w\|_2/\|y\|_2$ in (8.6.3), which has a maximum of 2×10^{10} without scaling and 1.5 with scaling. Note however that scaling increases the condition numbers $\widehat{\kappa}_P(\alpha,\beta)$ by factors ranging from 1.2 to 173.

Our third problem is a standard damped mass-spring system, as described in [83, Sec. 3.9]. The matrix $A = I$, B is tridiagonal with super and subdiagonal elements all -64 and diagonal $128, 192, 192, \ldots, 192$, and C is tridiagonal with super and subdiagonal elements all -1 and diagonal $2, 3, \ldots, 3$. Here, $\rho = 320$. The eigenvalues are all negative, with 50 eigenvalues of large modulus ranging from -320 to -6.4 and 50 small modulus eigenvalues approximately -1.5×10^{-2}. Figures 8.7.5 and 8.7.6 show the results for $v = e_1$ and $v = e_2$, respectively. Our theory suggests that for the eigenvalues of large modulus the linearization with $v = e_1$ will have nearly optimal conditioning, while for eigenvalues of small modulus the linearization with $v = e_2$ will be nearly optimal. This behaviour is seen very clearly in the figures, with a sharp change in condition number at the three order of magnitude jump in the eigenvalues. This example also clearly displays non-optimal conditioning of the first companion linearization for small eigenvalues: for the 50 eigenvalues of small modulus, $\widehat{\kappa}_{C_1}(\alpha,\beta)$ exceeds $\widehat{\kappa}_P(\alpha,\beta)$ and $\widehat{\kappa}_L(\alpha,\beta\,; e_2)$ by a factor at least 10^3, and again this is accurately reflected in the bounds (8.6.3). For this problem, scaling has essentially no effect on the two $\mathbb{DL}(Q)$ linearizations, but for the companion linearization it increases the condition number for the large eigenvalues and decreases it for the small eigenvalues, with the result that all the condition numbers lie between 3.6 and 13.

Finally, we briefly describe an example that emphasizes the importance in our analysis that the pencil $L \in \mathbb{DL}(P)$ be a linearization of P. The problem is a quadratic polynomial of dimension 107 arising from an acoustical model of a speaker box [45]. After scaling, $\rho = 1$. The computed eigenvalues from the companion form have moduli of order 1, except for two eigenvalues with moduli of order 10^{-5}. We found the pencil with $v = e_2$ to have eigenvalue condition numbers of the same order of magnitude as those of Q (namely from 10^6 to 10^{13})—as predicted by the theory. But for $v = e_1$ the conditioning of L was orders of magnitude worse than that of Q for *every* eigenvalue, not just the small ones, which at first sight appears to contradict the theory. The explanation is that this problem has a singular A_0 and hence a zero eigenvalue; L is therefore not a linearization for $v = e_1$, as we noted earlier (see (8.3.11), and the first sentence of the proof of Theorem 8.3.2). In fact, since $L \in \mathbb{DL}(P)$ for $v = e_1$ is not a linearization, it must by Theorem 2.2.3 be a singular pencil. This example is therefore entirely consistent with the theory.

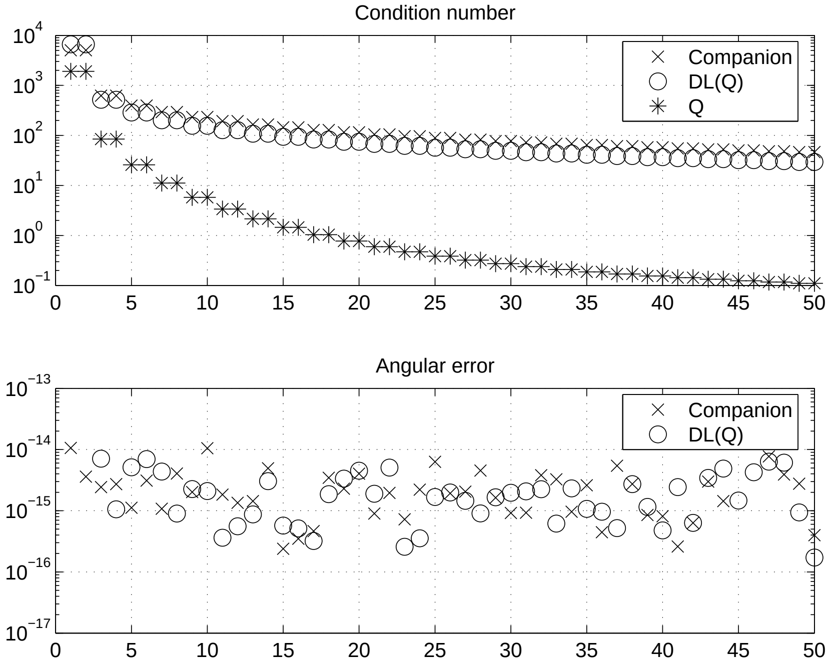


Figure 8.7.1: Wave problem, unscaled; $v = e_1$, $\rho = 625$.

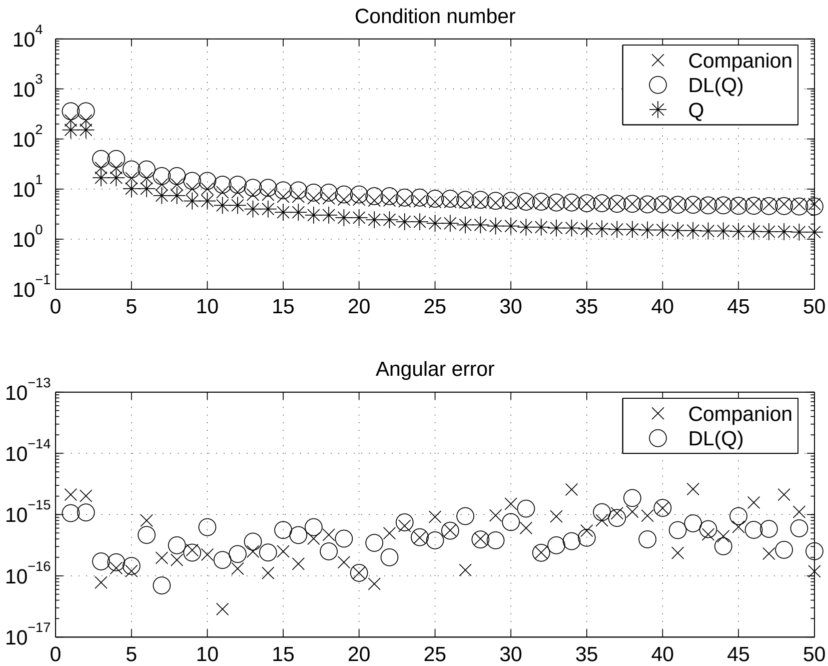


Figure 8.7.2: Wave problem, scaled; $v = e_2$, $\rho = 1$.

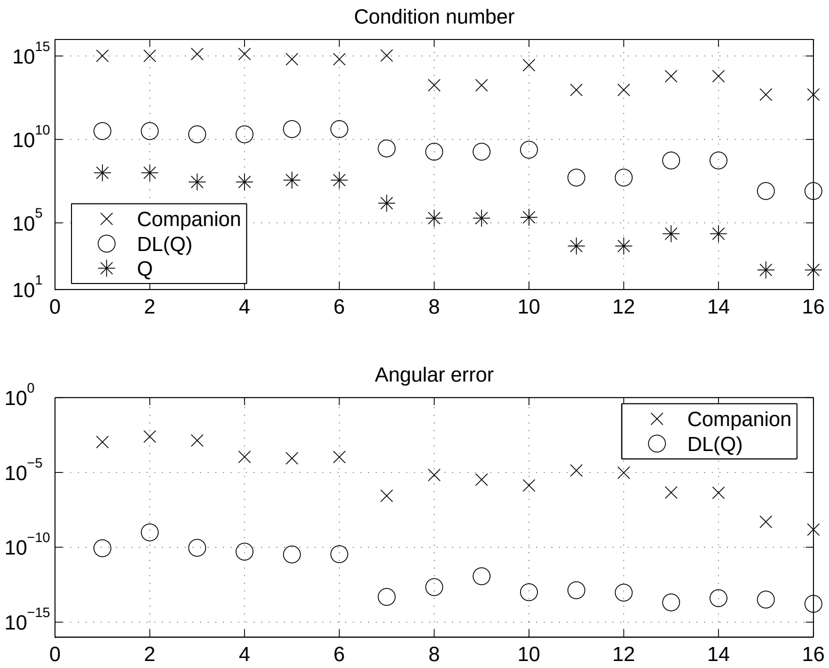


Figure 8.7.3: Nuclear power plant problem, unscaled; $v = e_1$, $\rho = 7 \times 10^4$.

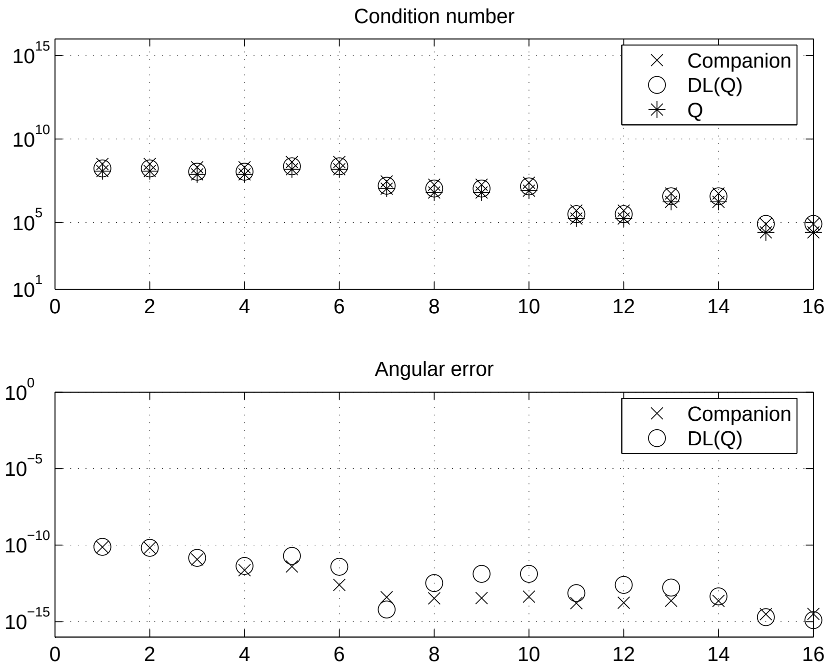


Figure 8.7.4: Nuclear power plant problem, scaled; $v = e_2$, $\rho = 1$.

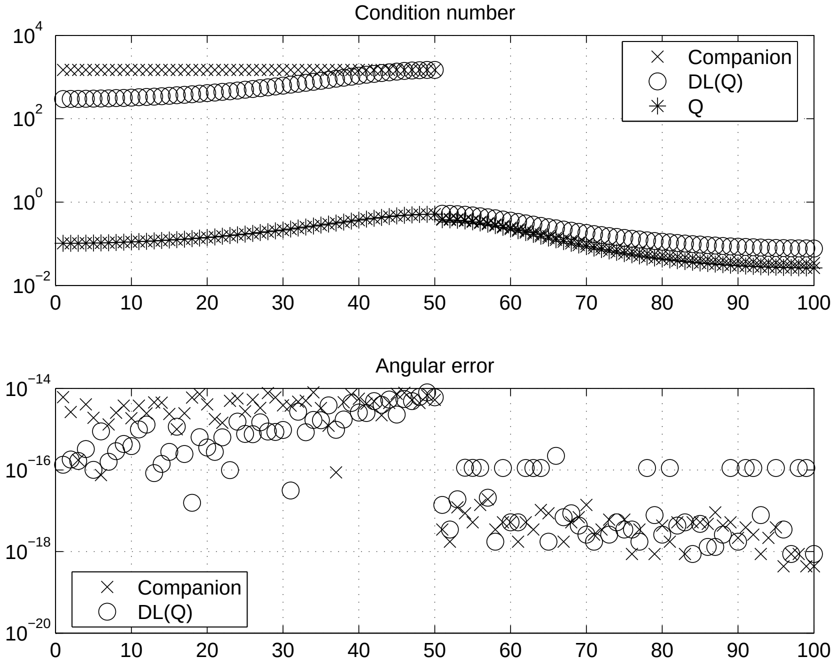


Figure 8.7.5: Damped mass-spring system, unscaled; $v = e_1$, $\rho = 320$.

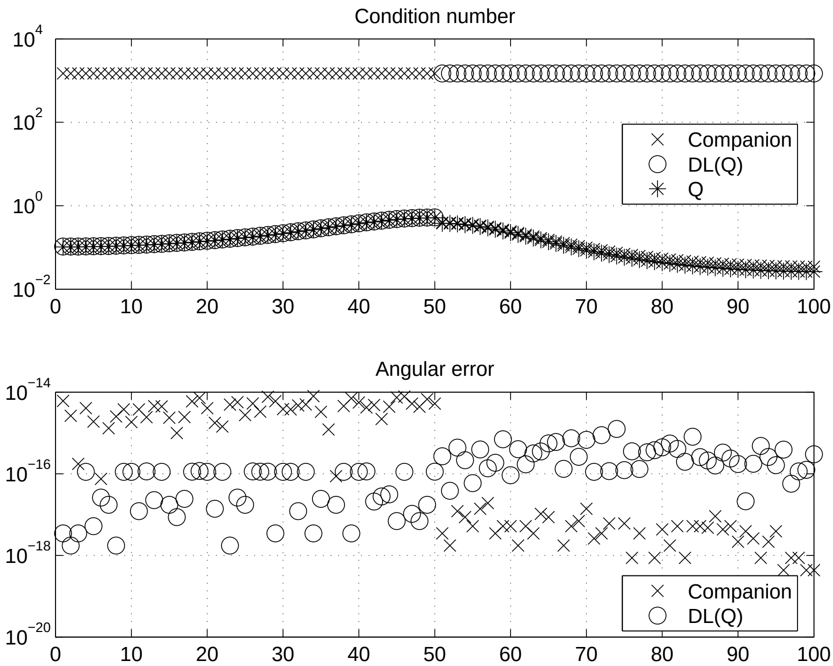


Figure 8.7.6: Damped mass-spring system, unscaled; $v = e_2$, $\rho = 320$.

8.8 Addendum — Some Useful Bounds

The following result is needed in the proofs of Theorems 8.3.3 and 8.3.5.

Proposition 8.8.1. *Consider the functions*

$$\begin{aligned} f_1(x) &= \frac{(1+x^2)(1+x^2+x^4+\cdots+x^{2(k-1)})}{1+x^{2k}}, \\ f_2(x) &= \frac{(1+x^2)(1+x^2+x^4+\cdots+x^{2(k-1)})}{1+x^2+x^4+\cdots+x^{2k}}, \\ f_3(x) &= \frac{(1+x)^2(1+x^2+x^4+\cdots+x^{2(k-1)})}{(1+x^k)^2}, \\ f_4(x) &= \frac{(1+x)^2(1+x^2+x^4+\cdots+x^{2(k-1)})}{(1+x+x^2+\cdots+x^k)^2}, \end{aligned}$$

for any $k \geq 2$. The functions f_1, f_2, f_3, and f_4 are all unimodal on the interval $[0, \infty)$, with a unique interior extreme point at $x = 1$ and another extreme point at $x = 0$. In particular, we have the following sharp bounds:

$$1 \leq f_1(x) \leq k, \tag{8.8.1}$$

$$1 \leq f_2(x) \leq \frac{2k}{k+1}, \tag{8.8.2}$$

$$1 \leq f_3(x) \leq k, \tag{8.8.3}$$

$$\frac{4k}{(k+1)^2} \leq f_4(x) \leq 1. \tag{8.8.4}$$

As a preliminary to proving these bounds, first observe that each of these four functions has the "reciprocal symmetry" property:

$$f(1/x) = f(x). \tag{8.8.5}$$

The qualitative behavior of such functions is somewhat constrained, as illustrated by the following lemma.

Lemma 8.8.2 (Unimodality of Reciprocally Symmetric Functions). *Suppose f is a differentiable function satisfying the property $f(1/x) = f(x)$ on the interval $(0, \infty)$. Then $x = 1$ must be a critical point for f. Furthermore, if f' has constant sign on either the interval $(0, 1)$ or the interval $(1, \infty)$, then f is unimodal on $(0, \infty)$ with a unique extreme point at $x = 1$.*

Proof. A simple calculation shows that $f(1/x) = f(x) \Rightarrow f'(1/x) \cdot (-1/x^2) = f'(x)$, and so

$$f'(1/x) = -x^2 f'(x). \tag{8.8.6}$$

Evaluating at $x = 1$ shows that $f'(1) = -f'(1)$, hence $f'(1) = 0$, so $x = 1$ is a critical point. The relation (8.8.6) also shows that f' has constant sign on $(0, 1)$ if and only if f' has constant and *opposite* sign on $(1, \infty)$, thus implying the unimodality of f and the existence of a unique extreme point at $x = 1$. ∎

A second more specialized result will also be needed to help establish the bounds in Proposition 8.8.1.

Lemma 8.8.3. *For any $k \geq 2$, the polynomial $p(x) = -x^{2k} + kx^{k+1} - kx^{k-1} + 1$ has exactly three positive roots, all at $x = 1$.*

Proof. Since there are three sign changes in the coefficients of p, Descartes' Rule of Signs says that $p(x)$ has either one or three positive roots. Clearly $p(1) = 0$, so $x = 1$ is one of these roots. But a straightforward calculation shows that $p'(1) = p''(1) = 0$, so $x = 1$ is a root of at least multiplicity three. This accounts for all possible positive roots that p could possibly have. $\square$

As final preparation for the proof of Proposition 8.8.1, note the following alternative expressions for the functions f_1 through f_4, found using the standard geometric series formula and valid for all $x \neq \pm 1$.

$$\begin{aligned}
f_1(x) &= \frac{1+x^2}{1-x^2} \cdot \frac{1-x^{2k}}{1+x^{2k}}\,, \\
f_2(x) &= 1 + \left[\frac{(x^2)^k - x^2}{(x^2)^{k+1} - 1}\right], \\
f_3(x) &= \frac{1+x}{1-x} \cdot \frac{1-x^k}{1+x^k}\,, \\
f_4(x) &= 1 - \left[\frac{x^k - x}{x^{k+1} - 1}\right]^2 .
\end{aligned}$$

These formulas not only make it easier to compute derivatives, but also reveal close connections between the functions that might have otherwise remained obscure. For example, we see that

$$f_1(x) = f_3(x^2) \tag{8.8.7}$$

on the interval $[0, \infty)$, a relation that is not at all evident from the original formulas.

Proof. [**of Proposition** 8.8.1] We begin with an analysis of f_3. Straightforward calculation shows that

$$f_3'(x) = \frac{2p(x)}{\big[(1-x)(1+x^k)\big]^2}\,,$$

where $p(x)$ is as in Lemma 8.8.3. Thus the only positive roots of either the numerator or denominator of f_3' are at $x = 1$, so f_3' has constant sign on the interval $(0, 1)$. Therefore f_3 is unimodal on $(0, \infty)$ by Lemma 8.8.2; this unimodality extends by continuity to the closed interval $[0, \infty)$. The bounds (8.8.3) are then obtained upon evaluating at $x = 0$ and $x = 1$. The unimodality of f_1 now follows from that of f_3 because of the relation (8.8.7); the bounds (8.8.1) again come from evaluating at $x = 0$ and $x = 1$.

The unimodality of f_2 and f_4 are obtained in a similar manner, using the derivative formulas

$$f_2'(x) = \frac{2xp(x^2)}{\big[(x^2)^{k+1} - 1\big]^2} \qquad \text{and} \qquad f_4'(x) = -2\left[\frac{x^k - x}{x^{k+1} - 1}\right] \cdot \frac{p(x)}{\big(x^{k+1} - 1\big)^2}\,,$$

where $p(x)$ is again the polynomial in Lemma 8.8.3. $\square$

Bibliography

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen, *LAPACK Users' Guide*, SIAM, Philadelphia, PA, second ed., 1995.

[2] A. L. Andrew, K.-W. E. Chu, and P. Lancaster, *Derivatives of eigenvalues and eigenvectors of matrix functions*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 903–926.

[3] E. N. Antoniou and S. Vologiannidis, *A new family of companion forms of polynomial matrices*, Electron. J. Linear Algebra, 11 (2004), pp. 78–87.

[4] T. Apel, V. Mehrmann, and D. Watkins, *Structured eigenvalue methods for the computation of corner singularities in 3D anisotropic elastic structures.*, Comp. Meth. Appl. Mech. Eng., 191 (2002), pp. 4459–4473.

[5] ——, *Numerical solution of large scale structured polynomial eigenvalue problems*, in Foundations of Computational Mathematics, Springer Verlag, 2003.

[6] Zhaojun Bai, James W. Demmel, Jack J. Dongarra, Axel Ruhe, and Henk A. van der Vorst, eds., *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[7] David Bellos, *Georges Perec: A Life in Words*, Harvill Press, London, 1995.

[8] Peter Benner, Ralph Byers, Volker Mehrmann, and Hongguo Xu, *Numerical computation of deflating subspaces of skew Hamiltonian/Hamiltonian pencils*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 165–190.

[9] Howard Bergerson, *Palindromes and Anagrams*, Dover Publications, New York, 1973.

[10] Dario A. Bini, Luca Gemignani, and Françoise Tisseur, *The Ehrlich-Aberth method for the nonsymmetric tridiagonal eigenvalue problem*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 153–175.

[11] M. A. Brebner and J. Grad, *Eigenvalues of $Ax = \lambda Bx$ for real symmetric matrices A and B computed by reduction to a pseudosymmetric form on the HR process*, Linear Algebra Appl., 43 (1982), pp. 99–118.

[12] T. J. Bridges and P. J. Morris, *Differential eigenvalue problems in which the parameter appears nonlinearly*, J. Comput. Phys., 55 (1984), pp. 437–460.

[13] Bill Bryson, *The Mother Tongue: English & How It Got That Way*, William Morrow and Co., New York, 1990.

[14] Angelika Bunse-Gerstner, *An analysis of the HR algorithm for computing the eigenvalues of a matrix*, Linear Algebra Appl., 35 (1981), pp. 155–173.

[15] Stephen J. Chism, *From A to Zotamorf: The Dictionary of Palindromes*, Word Ways Press, Morristown, NJ, 1992.

[16] D. Chu, X. Liu, and V. Mehrmann, *A numerically backwards stable method for computing the Hamiltonian Schur form*, Preprint 24-2004, Institut für Mathematik, TU Berlin, D-10623 Berlin, FRG, 2004. `www.math.tu-berlin.de/preprints/`.

[17] J.-P. Dedieu, *Condition operators, condition numbers, and condition number theorem for the generalized eigenvalue problem*, Linear Algebra Appl., 263 (1997), pp. 1–24.

[18] J.-P. Dedieu and F. Tisseur, *Perturbation theory for homogeneous polynomial eigenvalue problems*, Linear Algebra Appl., 358 (2003), pp. 71–94.

[19] Michael Donner, *I Love Me, Vol. I: S. Wordrow's Palindrome Encyclopedia*, Algonquin Books, Chapel Hill, N.C., 1996.

[20] Willard R. Espy, *Another Almanac of Words at Play*, Clarkson N. Potter, Inc., New York, 1980.

[21] ——, *The Word's Gotten Out*, Clarkson N. Potter, Inc., New York, 1989.

[22] E. Eich-Soellner and C. Führer, *Numerical Methods in Multibody Systems*, B. G. Teubner, Stuttgart, Germany, 1998.

[23] H.-Y. Fan, W.-W. Lin, and P. Van Dooren, *Normwise scaling of second order polynomial matrices*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 252–256.

[24] Heike Fassbender, *Symplectic Methods for the Symplectic Eigenvalue Problem*, Kluwer Academic, New York, 2000.

[25] Miroslav Fiedler, *A note on companion matrices*, Linear Algebra Appl., 372 (2003), pp. 325–331.

[26] P. Freitas, M. Grinfield, and P. Knight, *Stability of finite-dimensional systems with indefinite damping*, Advances in Mathematical Science and Applications, 17 (1997), pp. 435–446.

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

[28] F. R. Gantmacher, *Theory of Matrices*, vol. 2, Chelsea, New York, 1959.

[29] Seamus Garvey, Uwe Prells, Michael I. Friswell, and Zheng Chen, *General isospectral flows for linear dynamic systems*, Linear Algebra Appl., 385 (2004), pp. 335–368.

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

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

[32] ——, *Matrices and Indefinite Scalar Products*, Birkhauser, Basel, 1983.

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

[34] Nicholas J. Higham, D. Steven Mackey, Niloufer Mackey, and Françoise Tisseur, *Symmetric linearizations for matrix polynomials*, MIMS EPrint 2005.25, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Jan. 2006.

[35] Nicholas J. Higham, D. Steven Mackey, and Françoise Tisseur, *The conditioning of linearizations of matrix polynomials*, MIMS EPrint 2005.23, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Nov. 2005. To appear in SIAM J. Matrix Anal. Appl.

[36] Nicholas J. Higham, Françoise Tisseur, and Paul M. Van Dooren, *Detecting a definite Hermitian pair and a hyperbolic or elliptic quadratic eigenvalue problem, and associated nearness problems*, Linear Algebra Appl., 351–352 (2002), pp. 455–474.

[37] A. Hilliges, *Numerische Lösung von Quadratischen Eigenwertproblemen mit Anwendung in der Schienendynamik*, Diplomarbeit, TU Berlin, Inst. f. Mathematik, 2004.

[38] A. Hilliges, C. Mehl, and V. Mehrmann, *On the solution of palindromic eigenvalue problems*, in Proceedings of the 4th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS), Jyväskylä, Finland, 2004. CD-ROM.

[39] Peter Hilton, *Personal communication*, Oct 2004. See also Miscellanea, *Inventing a Palindrome*, College Math. J., 30, Nov 1999, p. 422.

[40] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[41] Tsung-Min Hwang, Wen-Wei Lin, Jinn-Liang Liu, and Weichung Wang, *Jacobi-Davidson methods for cubic eigenvalue problems*, Numer. Linear Algebra Appl., 12 (2005), pp. 605–624.

[42] Ilse C. F. Ipsen, *Accurate eigenvalues for fast trains*, SIAM News, 37(9), Nov 2004, pp. 1–2.

[43] T. Itoh, *Damped vibration mode superposition method for dynamic response analysis*, Earthquake Engrg. Struct. Dyn., 2 (1973), pp. 47–57.

[44] Scott Kim, *Inversions: A Catalog of Calligraphic Cartwheels*, W.H. Freeman, 1989.

[45] T. R. Kowalski, *Extracting a Few Eigenpairs of Symmetric Indefinite Matrix Pencils*, PhD thesis, Department of Mathematics, University of Kentucky, Lexington, KY 40506, USA, 2000.

[46] Peter Lancaster, *Symmetric transformations of the companion matrix*, NABLA: Bull. Malayan Math. Soc., 8 (1961), pp. 146–148.

[47] ——, *Lambda-Matrices and Vibrating Systems*, Pergamon Press, Oxford, 1966. Reprinted by Dover, New York, 2002.

[48] ——, *Quadratic eigenvalue problems*, Linear Algebra Appl., 150 (1991), pp. 499–506.

[49] P. Lancaster and P. Psarrakos, *A note on weak and strong linearizations of regular matrix polynomials*, Numerical Analysis Report No. 470, Manchester Centre for Computational Mathematics, Manchester, UK, June 2005.

[50] Peter Lancaster and Leiba Rodman, *The Algebraic Riccati Equation*, Oxford University Press, Oxford, 1995.

[51] Peter Lancaster and Miron Tismenetsky, *The Theory of Matrices*, Academic Press, London, 2nd ed., 1985.

[52] John Langdon, *Wordplay: The Philosophy, Art, and Science of Ambigrams*, Bantam Press, London, 2005.

[53] Richard Lederer, *The Word Circus: A Letter-Perfect Book*, Merriam-Webster, Inc., Springfield, Mass., 1998.

[54] D. Leguillon, *Computation of 3d-singularities in elasticity*, in Boundary value problems and integral equations in nonsmooth domains, M. C. et al., ed., vol. 167 of Lect. Notes Pure Appl. Math., New York, 1995, Marcel Dekker, pp. 161–170. Proceedings of the conference, held at the CIRM, Luminy, France, May 3-7, 1993.

[55] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, 1998.

[56] D. Steven Mackey, *The characteristic polynomial of a partitioned matrix*, in Linear Algebra Gems, D. Carlson et al., ed., MAA, Washington, DC, 2002, pp. 13–14. MAA Notes # 59.

[57] D. Steven Mackey, Niloufer Mackey, Christian Mehl, and Volker Mehrmann, *Structured polynomial eigenvalue problems: Good vibrations from good linearizations*, MIMS EPrint 2006.38, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Mar. 2006. (See also the earlier version "Palindromic polynomial eigenvalue problems", Apr. 2005.) Submitted to SIAM J. Matrix Anal. Appl.

[58] ——, *Vector spaces of linearizations for matrix polynomials*, MIMS EPrint 2005.26, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Dec. 2005. To appear in SIAM J. Matrix Anal. Appl.

[59] D. Steven Mackey, Niloufer Mackey, and Françoise Tisseur, *Structured factorizations in scalar product spaces*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 821–850.

[60] *MATLAB, Version 5.* The MathWorks, inc., 24 Prime Park Way, Natick, MA 01760-1500, USA, 1996.

[61] Volker Mehrmann, *The Autonomous Linear Quadratic Control Problem, Theory and Numerical Solution*, no. 163 in Lecture Notes in Control and Information Sciences, Springer-Verlag, Heidelberg, July 1991.

[62] —— , *A step toward a unified treatment of continuous and discrete time control problems*, Linear Algebra Appl., 241–243 (1996), pp. 749–779.

[63] V. Mehrmann and C. Shi, *Analysis of higher order linear differential-algebraic systems*, Preprint 17-2004, Institut für Mathematik, TU Berlin, D-10623 Berlin, FRG, 2004. `www.math.tu-berlin.de/preprints/` .

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

[65] —— , *Polynomial eigenvalue problems with Hamiltonian structure*, Electron. Trans. Numer. Anal., 13 (2002), pp. 106–118.

[66] O.V. Michaelsen, *Never Even or Odd: Palindromes, Anagrams, & Other Tricks Words Can Do*, Sterling Publishing Co., 2005.

[67] Warren F. Motte, Jr., *Oulipo: A Primer of Potential Literature*, University of Nebraska Press, 1986.

[68] Beresford N. Parlett and H. C. Chen, *Use of indefinite pencils for computing damped natural modes*, Linear Algebra Appl., 140 (1990), pp. 53–88.

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

[70] R. Reemtsen and S. Görner, *Numerical methods for semi-infinite programming: A survey*, in Semi-Infinite Programming, R. Reemtsen and J.-J. Rückmann, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 195–275.

[71] Steve Rozen, Helen Skaletsky, ... (5 authors) ..., and David Page, *Abundant gene conversion between arms of palindromes in human and ape Y chromosomes*, Nature, 423 (19 June 2003), pp. 873-876.

[72] Heinz Rutishauser, *Solution of eigenvalue problems with the LR-transformation*, in Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues, no. 49 in Applied Mathematics Series, National Bureau of Standards, United States Department of Commerce, Washington, D. C., 1958, pp. 47–81.

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

[74] F. Schmidt, T. Friese, L. Zschiedrich, and P. Deuflhard, *Adaptive multigrid methods for the vectorial Maxwell eigenvalue problem for optical waveguide design*, in Mathematics. Key Technology for the Future, W. Jäger and H.-J. Krebs, eds., Springer-Verlag, 2003, pp. 279–292.

[75] Christian Schröder, *A canonical form for palindromic pencils and palindromic factorizations*, Technical Report 316, DFG Research Center Matheon, *Mathematics for key technologies* in Berlin, TU Berlin, Str. des 17. Juni 136, D-10623 Berlin, Germany, 2006.
`www.matheon.de/research/list_preprints.asp` .

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

[77] Helen Skaletsky, ... (38 authors) ..., and David Page, *The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes*, Nature, 423 (19 June 2003), pp. 825-837.

[78] Hans Stetter, *Numerical Polynomial Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.

[79] G. W. Stewart and Ji-guang Sun, *Matrix Perturbation Theory*, Academic Press, New York, 1990.

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

[81] ——, *Tridiagonal-diagonal reduction of symmetric indefinite pairs*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 215–232.

[82] Françoise Tisseur and Nicholas J. Higham, *Structured pseudospectra for polynomial eigenvalue problems, with applications*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 187–208.

[83] Françoise Tisseur and Karl Meerbergen, *The quadratic eigenvalue problem*, SIAM Rev., 43 (2001), pp. 235–286.

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

[85] Krešimir Veselić, *A Jacobi eigenreduction algorithm for definite matrix pairs*, Numer. Math., 64 (1993), pp. 241–269.

[86] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, 1965.

[87] Huntington F. Willard, *Tales of the Y chromosome*, Nature, 423 (19 June 2003), pp. 810-812.

[88] Hongguo Xu, *Equivalence between generalized symplectic pencil and skew-Hermitian/Hermitian pencil*, Technical Report 175, DFG Research Center Matheon, *Mathematics for key technologies* in Berlin, TU Berlin, Str. des 17. Juni 136, D-10623 Berlin, Germany, 2005.
`www.matheon.de/research/list_preprints.asp` .

[89] Sabine Zaglmayr, *Eigenvalue problems in SAW-filter simulations*, Diplomarbeit, Inst. of Comp. Mathematics, J. Kepler Univ. Linz, Austria, 2002.