

Canonical Forms for Hermitian Matrix Pairs under Strict Equivalence and Congruence

Peter Lancaster
Department of Mathematics and Statistics
University of Calgary
Calgary, AB T2N 1N4, Canada*
lancaste@math.ucalgary.ca

Leiba Rodman
Department of Mathematics
College of William and Mary
Williamsburg, VA,
USA*
lxrodm@math.wm.edu

Abstract

After a brief historical review and an account of the canonical forms attributed to Jordan and Kronecker, a systematic development is made of the simultaneous reduction of pairs of quadratic forms over the complex numbers and over the reals. These reductions are by strict equivalence and by congruence and essentially complete proofs are presented. Some closely related results which can be derived from the canonical forms are also included. They concern simultaneous diagonalization, a new criterion for the existence of positive definite linear combinations of a pair of hermitian matrices, and the canonical structures of matrices which are selfadjoint in an indefinite inner product.

Key words: canonical forms, matrix pairs, strict equivalence, congruence.

Mathematics Subject Classification 2000: 15A22, 15A21, 15A63.

*The research of Peter Lancaster was supported in part by the Natural Sciences and Engineering Research Council of Canada and a Senior Fellowship of the University of Manchester, UK. The research of Leiba Rodman was supported in part by a Summer Research Grant of the College of William and Mary.

1 Introduction

The theory of bilinear and quadratic forms occupies a central position in mathematics and its applications. Knowledge of the simplifications made possible by transformations of coordinates are frequently of great importance in geometry, algebra, statistics, function theory, and many other areas. Here, attention is confined to forms over either the real or complex number fields (\mathbb{R} or \mathbb{C}). The complete reduction of a single form was well-understood in the late nineteenth century (the paper of C. Jordan, after which the Jordan canonical form takes its name, was published in 1870). But many fields of research required an understanding of the *simultaneous* reduction of two forms. Furthermore, these forms may have properties of symmetry or skew-symmetry in each of the three possible combinations; they may also be definite or indefinite, and they may be singular or regular. To illustrate, the simultaneous reduction of two *real symmetric* quadratic forms representing kinetic and potential energies is central in the theory of vibrations and, in one form or another was probably known to Lagrange in the eighteenth century and effectively used by Rayleigh in the first edition of his “Theory of Sound” in 1877 (see [48]).

Thus, by the late nineteenth century, the importance of the simultaneous reduction of more complicated pairs of forms (admitting singular forms and other symmetries) was recognized and several important contributions were made by Weierstrass (1868) (see [62]), by Kronecker through the 1870’s (see [27]), and Sylvester through the 1880’s (see Wedderburn, [57], where there is a bibliography of 549 items covering the period 1853-1933). Further analyses were completed through the early twentieth century and a summary can be found in the Historical Notes to Chapter 9 of the work of Turnbull and Aitken [55], as well as the classic survey of MacDuffee [36] of the same era.

In the second half of the twentieth century developments in operator theory, systems theory and control, and signal processing (among others) demanded detailed understanding of problems of this general character. In particular, the importance of Hamiltonian and symplectic symmetries became apparent. Seminal work was done by Pontryagin in 1940 [43] and by M. G. Krein in 1955 [24]. There were subsequent applications to damped oscillatory systems and to factorization of matrix and operator functions by Krein and Langer [26], Langer [34], and Gohberg et al. [14]. More recent contributions in this direction with emphasis on numerically stable algorithms can be found in works of Benner et al. [5], Lin et al. [35], Mehl et al [39], and Mehrmann and Xu [40].

As the simultaneous reductions of pairs of forms arose in diverse parts of mathematics it was, perhaps, inevitable that results would be scattered throughout the literature with some duplications. Accordingly, R. C. Thompson undertook a review of the field in 1973 leading to the work of [51], although the review itself was not published until 1991 (see [52]). This is an important contribution containing a reference list of 225 items. Here, a second review is provided, and one may well ask why!

Undoubtedly, the authors have benefited greatly from the works of Thompson, but

there is still a strong case for a systematic treatment of the theory and presentation of results which will facilitate easy reference by the many and varied users. Also, where sketch-proofs have been provided previously, more detail is to be given here, and the proofs are geared more to the particular cases of complex and real matrices. Thus, the exposition is to be as self-contained as is reasonably possible, starting from the more familiar canonical forms of Jordan (Theorems 2.1 and 2.2) and Kronecker (Theorems 3.2 and 3.3). These results are stated without proofs, and they also serve to establish some (often intricate) notations which will be used throughout this review.

Since there is a large body of work here, this first paper is devoted to problems concerning reductions of two hermitian matrices. Thus, Theorems 5.1 and 6.1 are the central results concerning reduction of two complex hermitian matrices. The line of proof used here seems to be new, although it is a natural line of argument given the Kronecker forms. Furthermore, these results are not easy to find elsewhere in the literature. Theorems 9.1 and 9.2 are the corresponding results concerning pairs of real symmetric matrices.

Sections 10, 11, and 12 contain some important closely related results which can be derived from the canonical forms. They concern simultaneous diagonalization (Section 10), a new criterion for the existence of positive definite (or semidefinite) linear combinations of a pair of hermitian matrices (Section 11), and the canonical structures of matrices which are selfadjoint in an indefinite inner product (Section 12).

A similar analysis for pairs of forms with at least one of them skew-symmetric will be the subject of a second paper (see [30]).

Let us turn now to questions of notation and terminology. Basic ideas concerning pairs of matrices can be formulated in the context of so-called matrix pencils: $A + \lambda B$ where $A, B \in \mathbb{C}^{m \times n}$ and λ is a scalar complex parameter. Some basic definitions are made here in the context of complex matrices. The reader can fill in the corresponding definitions over \mathbb{R} .

Matrices will frequently be treated as linear transformations on finite dimensional vector spaces over \mathbb{R} , or \mathbb{C} . Thus, the context is frequently either the real euclidean space of real n -tuple columns, \mathbb{R}^n over \mathbb{R} , or the space of complex n -tuple columns, \mathbb{C}^n over \mathbb{C} . The *standard inner product* in these spaces is defined by writing

$$(x, y) = \sum_{j=1}^n x_j \bar{y}_j.$$

Denote by $\mathbb{C}^{m \times n}$, resp., $\mathbb{R}^{m \times n}$ the space of $m \times n$ complex, resp., real, matrices. If $A \in \mathbb{C}^{n \times n}$ has entries a_{ij} , the matrix with corresponding entries \bar{a}_{ji} is denoted by A^* and is known as the *adjoint* of A , and may also be written as \bar{A}^T where the index T denotes transposition. A matrix $A \in \mathbb{C}^{n \times n}$ is hermitian (or self-adjoint) if $A^* = A$ and unitary if $A^*A = I_n$, the identity matrix in $\mathbb{C}^{n \times n}$. A matrix $A \in \mathbb{R}^{n \times n}$ is, of course, symmetric if $A^T = A$ and (real) orthogonal if $A^T A = I_n$. The $u \times v$ zero matrix will be denoted $0_{u \times v}$.

Thus, an $n \times n$ complex matrix is frequently identified with a linear transformation acting on \mathbb{C}^n in the usual way (and similarly for real matrices). With this understanding, and with the standard inner product, \mathbb{C}^n becomes a Hilbert space, and A^* is, indeed, the adjoint of A in the Hilbert space sense.

Matrices A and B from $\mathbb{C}^{n \times n}$ are said to be *similar* if there is a nonsingular matrix T such that $A = TBT^{-1}$, and it is well-known that this matrix transformation corresponds to a change of basis in \mathbb{C}^n .

The matrix pencils $A_1 + \lambda B_1$ and $A_2 + \lambda B_2$ (or the pairs (A_1, B_1) and (A_2, B_2)) in $\mathbb{C}^{m \times n}$ are said to be *strictly equivalent* if there exist nonsingular P and Q such that

$$P(A_1 + \lambda B_1)Q = A_2 + \lambda B_2, \quad \text{for all } \lambda \in \mathbb{C}. \quad (1.1)$$

Thus, such a transformation corresponds to changes of bases in the domain and range spaces of the pencil.

The more restrictive concept of congruence applies to pencils of *square* matrices. Thus, pencils $A_1 + \lambda B_1$ and $A_2 + \lambda B_2$ in $\mathbb{C}^{n \times n}$ are said to be *congruent* if there exists a nonsingular P such that

$$P(A_1 + \lambda B_1)P^* = A_2 + \lambda B_2, \quad \text{for all } \lambda \in \mathbb{C}.$$

A classical result of this kind says that if A_1 and B_1 are hermitian and B_1 is positive definite, then there is a P for which $B_2 = I_n$ and A_2 is real and diagonal. This can be achieved by an appropriate choice of basis in the domain space. The generalization of this result to admit any hermitian A_1 and B_1 is an important part of the theory to be presented here.

2 The Jordan form

The complex vector space $\mathbb{C}^{n \times n}$ can be subdivided into disjoint equivalence classes of similar matrices. Each equivalence class is characterized by a unique matrix in canonical form. Conventions vary to some degree on the specification of this canonical form. It forms a building block for other forms developed in the sequel.

First define a *Jordan block* of size m by

$$J_m(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{bmatrix} \in \mathbb{C}^{m \times m}.$$

Then the fundamental theorem specifying the Jordan canonical form is as follows:

Theorem 2.1 For any $A \in \mathbb{C}^{n \times n}$ there is a block-diagonal matrix $J \in \mathbb{C}^{n \times n}$ of the form

$$J = \begin{bmatrix} J_{m_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{m_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & J_{m_r}(\lambda_r) \end{bmatrix} \quad (2.1)$$

which is similar to A .

Moreover, the matrix J is uniquely determined by A up to permutation of the diagonal blocks, i.e. if A is also similar to a matrix

$$\begin{bmatrix} J_{n_1}(\mu_1) & 0 & \cdots & 0 \\ 0 & J_{n_2}(\mu_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & J_{n_s}(\mu_s) \end{bmatrix},$$

then $r = s$ and the collection $\{J_{n_1}(\mu_1), J_{n_2}(\mu_2), \dots, J_{n_s}(\mu_s)\}$ (possibly with repeated elements) can be re-arranged so that $J_{n_k}(\mu_k) = J_{m_k}(\lambda_k)$ for $k = 1, 2, \dots, r$.

Proofs of this theorem can be found in many textbooks on linear algebra and the theory of matrices such as Finkbeiner [9], Gantmacher [12], Gohberg et al. [15], Horn and Johnson [19], Lancaster and Tismenetsky [31], Smith [50].

The following terminology is adopted: The matrix (2.1) is called the *Jordan form* of A (or possibly the *complex Jordan form* of A). The numbers $\lambda_1, \lambda_2, \dots, \lambda_r$ are the *eigenvalues* of A (not necessarily distinct), and the set of all eigenvalues is the *spectrum* of A , denoted by $\sigma(A)$.

For an eigenvalue λ_0 of A , the *geometric multiplicity*, $\gamma(\lambda_0)$ is the number of Jordan blocks in the Jordan form of A in which λ_0 appears as an eigenvalue (i.e. in which $\lambda_j = \lambda_0$). The geometric multiplicity of an eigenvalue can also be defined as the dimension of the corresponding eigenspace. The *algebraic multiplicity*, $\alpha(\lambda_0)$, of eigenvalue λ_0 is the sum of the sizes of Jordan blocks in which λ_0 appears as an eigenvalue. It is easily verified that

$$\gamma(\lambda_0) = \dim \text{Ker}(A - \lambda_0 I), \quad \alpha(\lambda_0) = \dim \text{Ker}(A - \lambda_0 I)^n. \quad (2.2)$$

There are, of course, analogous equivalence classes of *real* square matrices in $\mathbb{R}^{n \times n}$ generated by real similarity transformations. It turns out that the corresponding canonical forms are not unlike those above but, because of the presence of eigenvalues which may be real or non-real in conjugate pairs, the canonical forms are rather more complicated. Perhaps for this reason, they are generally less familiar.

Description of an appropriate canonical form requires the introduction of another class of matrices in standard form (which will account for the presence of non-real

eigenvalues). For real numbers, λ and $\mu \neq 0$ define the *real Jordan block* of even size, say $2m \times 2m$, by

$$J_{2m}(\lambda \pm i\mu) = \begin{bmatrix} \lambda & \mu & 1 & 0 & \cdots & 0 & 0 \\ -\mu & \lambda & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \mu & \cdots & 0 & 0 \\ 0 & 0 & -\mu & \lambda & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & & 0 & 1 \\ 0 & 0 & 0 & 0 & & \lambda & \mu \\ 0 & 0 & 0 & 0 & & -\mu & \lambda \end{bmatrix} \in \mathbb{R}^{2m \times 2m}.$$

Thus, there are m real 2×2 blocks on the main diagonal and 2×2 identity matrices making up the super-diagonal blocks. Clearly $J_{2m}(\lambda \pm i\mu)$ is a real matrix, and

$$\sigma(J_{2m}(\lambda \pm i\mu)) = \{\lambda + i\mu, \lambda - i\mu\}.$$

If X_1, \dots, X_p are matrices, we denote by $X_1 \oplus \dots \oplus X_p$ the block diagonal matrix with the diagonal blocks X_1, \dots, X_p (in that order). Then the Jordan form over \mathbb{R} is given by:

Theorem 2.2 *For any $A \in \mathbb{R}^{n \times n}$ there is a block-diagonal matrix $J \in \mathbb{R}^{n \times n}$ which is similar to A over the reals (i.e. there is an invertible $S \in \mathbb{R}^{n \times n}$ such that $A = S^{-1}JS$) and has the form*

$$J = J_{m_1}(\lambda_1) \oplus J_{m_2}(\lambda_2) \oplus \cdots \oplus J_{m_r}(\lambda_r) \oplus J_{2m_{r+1}}(\lambda_{r+1} \pm i\mu_{r+1}) \oplus \cdots \oplus J_{2m_q}(\lambda_q \pm i\mu_q), \quad (2.3)$$

where the λ_j are real and the μ_j are real and positive.

Moreover, the matrix J of (2.3) is uniquely determined by A up to permutation of the diagonal blocks.

Naturally, the matrix J of (2.3) is known as the *real Jordan form* of A . A complete proof of this theorem can be found in Gohberg et al, [17, Chapter 12]. (Another source is Shilov [49].)

For easy reference, it is convenient to record here a complex similarity transformation of the real standard matrix $J_{2m}(\lambda + i\mu)$ to its *complex* Jordan form. First define an invertible $2m \times 2m$ matrix

$$C_{2m} = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \oplus \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}.$$

The inverse is easily computed:

$$C_{2m}^{-1} = \frac{1}{2i} \left(\begin{bmatrix} i & -1 \\ i & 1 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} i & -1 \\ i & 1 \end{bmatrix} \right).$$

Use the unit coordinate vectors e_j , $j = 1, 2, \dots, 2m$, to define the permutation matrix

$$D_{2m} = [e_1 \ e_3 \ e_5 \ \dots \ e_{2m-1} \ e_2 \ e_4 \ \dots \ e_{2m}] \in \mathbb{R}^{2m \times 2m},$$

and observe that $D_{2m}^{-1} = D_{2m}^T$. Then a computation serves to establish the similarity

$$(C_{2m}D_{2m})^{-1}J_{2m}(\lambda \pm i\mu)(C_{2m}D_{2m}) = J_m(\lambda - i\mu) \oplus J_m(\lambda + i\mu). \quad (2.4)$$

This section is concluded with several useful formulas concerning similarity which can be verified by straightforward computation. Introduce the $m \times m$ real symmetric matrix

$$F_m = \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 \\ \vdots & & & 1 & 0 \\ \vdots & & & \vdots & \\ 0 & 1 & & \vdots & \\ 1 & 0 & \cdots & \cdots & 0 \end{bmatrix} = F_m^{-1}, \quad (2.5)$$

(with $F_1 = [1]$). As in [16], this matrix F_m is often called the *sip* matrix of size m (an abbreviation for standard involutory permutation). It is also known as the “reverse identity”. Clearly,

$$F_m J_m(\lambda) F_m = J_m(\lambda)^T, \quad (2.6)$$

and by taking complex conjugates,

$$F_m J_m(\bar{\lambda}) F_m = J_m(\lambda)^*. \quad (2.7)$$

The real analogue of (2.6) also holds:

$$F_{2m} J_{2m}(\lambda \pm i\mu) F_{2m} = J_{2m}(\lambda \pm i(-\mu))^T. \quad (2.8)$$

Note that $J_{2m}(\lambda \pm i\mu)$ is similar (over the reals) to $J_{2m}(\lambda \pm i(-\mu))$. Indeed,

$$\begin{bmatrix} F_2 & 0 & \cdots & 0 \\ 0 & F_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & F_2 \end{bmatrix} J_{2m}(\lambda \pm i\mu) \begin{bmatrix} F_2 & 0 & \cdots & 0 \\ 0 & F_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & F_2 \end{bmatrix} = J_{2m}(\lambda \pm i(-\mu)). \quad (2.9)$$

This allows one to replace the condition $\mu_j > 0$ of Theorem 2.2 by the condition that $\mu_j < 0$ when this is more convenient.

3 The Kronecker form

A canonical form for rectangular pencils $A + \lambda B \in \mathbb{C}^{m \times n}$ under strict equivalence is known as the *Kronecker form*. It incorporates some of the structures described above for

the Jordan canonical form. If $m = n$ in (1.1) and $B_1 = B_2 = I$, then strict equivalence reduces to a similarity between A_1 and A_2 . In this sense, the Kronecker form must generalize the Jordan form, and the re-appearance of Jordan blocks is plausible.

An $m \times n$ pencil $A + \lambda B$ is said to be *singular* if either $m \neq n$ or if $m = n$ and the determinant $\det(A + \lambda B)$ vanishes identically. The *rank* of such a pencil is the size of a largest minor which does not vanish identically. A pencil which is not singular is said to be *regular*. The *eigenvalues* are the (discrete) points λ_0 in $\mathbb{C} \cup \{\infty\}$ at which the rank of $A + \lambda_0 B$ is less than that of the pencil.

To describe the Kronecker form it is convenient to introduce some rectangular (and therefore singular) matrix pencils of special form. Define pencils in $\mathbb{C}^{\varepsilon \times (\varepsilon+1)}$ by

$$L_{\varepsilon \times (\varepsilon+1)} = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \end{bmatrix}.$$

In what follows, we denote by $0_{u \times v}$ the $u \times v$ zero matrix.

Here, and subsequently, the canonical form is presented in three or four parts:

1. The canonical blocks associated with the “singular structure” (and determining the *left and right indices* - see below).
2. The canonical blocks associated with the eigenvalue at infinity.
3. The canonical blocks associated with finite eigenvalues.

Where appropriate, Item 3 is broken down further to show:

- 3a. The blocks associated with finite real eigenvalues.
- 3b. The blocks associated with finite non-real eigenvalues.

Theorem 3.1 *Every pencil $A + \lambda B \in \mathbb{C}^{m \times n}$ is strictly equivalent to a matrix pencil with the block-diagonal form:*

$$\begin{aligned} 0_{u \times v} \oplus L_{\varepsilon_1 \times (\varepsilon_1+1)} \oplus \cdots \oplus L_{\varepsilon_p \times (\varepsilon_p+1)} \oplus L_{\eta_1 \times (\eta_1+1)}^T \oplus L_{\eta_q \times (\eta_q+1)}^T \oplus \\ (I_{k_1} + \lambda J_{k_1}(0)) \oplus \cdots \oplus (I_{k_r} + \lambda J_{k_r}(0)) \oplus \\ (\lambda I_{\ell_1} + J_{\ell_1}(\alpha_1)) \oplus \cdots \oplus (\lambda I_{\ell_s} + J_{\ell_s}(\alpha_s)), \end{aligned} \quad (3.1)$$

where $\varepsilon_1 \leq \dots \leq \varepsilon_p$; $\eta_1 \leq \dots \leq \eta_q$; $k_1 \leq \dots \leq k_r$; and $\ell_1 \leq \dots \leq \ell_s$ are positive integers.

Moreover, the integers ε_i , η_j , k_u are uniquely determined by the pair A , B , and the part

$$(\lambda I_{\ell_1} + J_{\ell_1}(\alpha_1)) \oplus \cdots \oplus (\lambda I_{\ell_s} + J_{\ell_s}(\alpha_s))$$

is uniquely determined by A and B up to a permutation of the diagonal blocks.

This block diagonal matrix is the *Kronecker form* for $A + \lambda B$. The proof is not reproduced here; it can be found in Gantmacher [12], or in the Appendix of Gohberg

et al. [17], for example. In this form, it may happen that some of the parameters u, v, p, q, r and s are zero. This means that the corresponding blocks are missing from (3.1). In particular, if $A + \lambda B$ is regular then the zero block, the L , and the L^T blocks do not appear. Thus, the first row of (3.1) may be described as “the singular part” of the Kronecker form and the last two rows as the “regular part”.

It will be useful to summarize other important features of the Kronecker form. If we define the *left kernel* of a matrix $Z \in \mathbb{C}^{m \times n}$ by

$$\text{Kel } Z = \{x \in \mathbb{C}^{1 \times m} : xZ = 0\}$$

then the parameters u and v of the first (zero) block can be identified as:

$$u = \dim \bigcap_{\lambda \in \mathbb{C}} \text{Kel}(A + \lambda B), \quad (3.2)$$

$$v = \dim \bigcap_{\lambda \in \mathbb{C}} \text{Ker}(A + \lambda B). \quad (3.3)$$

The integers $\varepsilon_1 \leq \dots \leq \varepsilon_p$ are called the *right indices* of the pencil $A + \lambda B$. They can be identified as the degrees of polynomial vector columns $x_1(\lambda), \dots, x_p(\lambda)$ forming a basis for the vector space of solutions $x(\lambda)$ of the equation $(A + \lambda B)x(\lambda) \equiv 0$.

Similarly, $\eta_1 \leq \dots \leq \eta_q$ are the *left indices* of $A + \lambda B$. They can be identified with the degrees of polynomial vector rows $y_1(\lambda), \dots, y_q(\lambda)$ forming a basis for the vector space of solutions $y(\lambda)$ of the polynomial equation $y(\lambda)(A + \lambda B) \equiv 0$.

The integers $k_1 \leq \dots \leq k_r$ are the *indices at infinity* of $A + \lambda B$ and, finally, the part

$$(\lambda I_{\ell_1} + J_{\ell_1}(\alpha_1)) \oplus \dots \oplus (\lambda I_{\ell_s} + J_{\ell_s}(\alpha_s)) \quad (3.4)$$

is the *Jordan part* of $A + \lambda B$. The numbers $-\alpha_1, \dots, -\alpha_s$ are said to be the *eigenvalues* of $A + \lambda B$, and the sizes of the blocks in (3.4) that correspond to a fixed eigenvalue $-\alpha$ are called the *indices* of $A + \lambda B$ corresponding to $-\alpha$. Obviously, if B is square and invertible, then the Kronecker form reduces to just the Jordan part.

As with the Jordan forms, the question of the analogue of the Kronecker form for real pencils strictly equivalent over the reals arises naturally (i.e. the case in which all the matrices appearing in (1.1) are real and $\lambda \in \mathbb{R}$). The corresponding canonical form can be formulated using the standard blocks formulated above.

Theorem 3.2 *Every pencil $A + \lambda B \in \mathbb{R}^{m \times n}$ is strictly equivalent over the reals to a matrix pencil with the block-diagonal form:*

$$\begin{aligned} 0_{u \times v} \oplus L_{\varepsilon_1 \times (\varepsilon_1 + 1)} \oplus \dots \oplus L_{\varepsilon_p \times (\varepsilon_p + 1)} \oplus L_{\eta_1 \times (\eta_1 + 1)}^T \oplus L_{\eta_q \times (\eta_q + 1)}^T \oplus \\ (I_{k_1} + \lambda J_{k_1}(0)) \oplus \dots \oplus (I_{k_r} + \lambda J_{k_r}(0)) \oplus \\ (\lambda I_{\ell_1} + J_{\ell_1}(\alpha_1)) \oplus \dots \oplus (\lambda I_{\ell_s} + J_{\ell_s}(\alpha_s)) \oplus \\ (\lambda I_{2j_1} + J_{2j_1}(\mu_1 \pm i\nu_1)) \oplus \dots \oplus (\lambda I_{2j_t} + J_{2j_t}(\mu_t \pm i\nu_t)) \end{aligned} \quad (3.5)$$

where $\varepsilon_1 \leq \dots \leq \varepsilon_p$; $\eta_1 \leq \dots \leq \eta_q$; $k_1 \leq \dots \leq k_r$ are positive integers, and α_j, μ_w, ν_w are real numbers with $\nu_1 > 0, \dots, \nu_t > 0$.

Moreover, the integers $\varepsilon_i, \eta_j, k_u$ are uniquely determined by the pair A, B , and each of the two parts

$$(\lambda I_{\ell_1} + J_{\ell_1}(\alpha_1)) \oplus \dots \oplus (\lambda I_{\ell_s} + J_{\ell_s}(\alpha_s))$$

and

$$(\lambda I_{2j_1} + J_{2j_1}(\mu_1 \pm i\nu_1)) \oplus \dots \oplus (\lambda I_{2j_t} + J_{2j_t}(\mu_t \pm i\nu_t))$$

is uniquely determined by A and B up to a permutation of the diagonal blocks.

The form (3.5) is the *real Kronecker form* of $A + \lambda B$. The (real) left indices, right indices, and the indices at infinity of $A + \lambda B$ are defined as in the complex case above. The two parts of the form displayed immediately above are, of course, the *real Jordan part* of $A + \lambda B$. The reader is referred to Gantmacher [12, Chapter XII] for the proof of this theorem.

The following corollary can be easily obtained by comparing real and complex canonical forms of real matrix pencils:

Corollary 3.3 *If two real matrix pencils $A + \lambda B$ and $A_1 + \lambda B_1$ are strictly equivalent over \mathbb{C} , then they are also strictly equivalent over \mathbb{R} .*

Note that this is a general phenomenon: if $F_1 \subseteq F_2$ are fields, and if two pencils $A + \lambda B$ and $A_1 + \lambda B_1$, with $A, B, A_1, B_1 \in F_1^{m \times n}$ are strictly equivalent over F_2 , then $A + \lambda B$ and $A_1 + \lambda B_1$ are also strictly equivalent over F_1 .

Uniqueness of the Kronecker form (up to certain permutations of blocks), and formula (2.6) yield:

Corollary 3.4 *A matrix pencil $A + \lambda B \in \mathbb{C}^{n \times n}$ is strictly equivalent to its transpose $A^T + \lambda B^T$ if and only if the right indices of $A + \lambda B$ coincide with its left indices.*

Notice that this corollary certainly holds if the pencil is regular, for then there are no indices at all.

Note also that canonical structures for real pairs (without symmetry) under strict equivalence, and those for symmetric pairs under congruence can be expected to have a lot in common; and this will be confirmed in what follows (compare Theorems 9.1 and 9.2). One reason for this is a result of Lancaster and Ye (Theorem 4 of [32]) showing that, for regular pairs, the reduction of nonsymmetric real pairs under strict equivalence is equivalent to the reduction of real symmetric pairs by congruence. Thus, if $A, B \in \mathbb{R}^{n \times n}$ and $A + \lambda B$ is regular, there is a nonsingular real matrix P such that PA and PB are symmetric.

4 Congruence of hermitian matrices

Some useful and well-known properties of hermitian matrices concerning congruence and inertia are collected in this section as a prologue to the more sophisticated results to follow.

Of course, matrices $H_1, H_2 \in \mathbb{C}^{n \times n}$ are said to be *congruent* if $H_1 = T^* H_2 T$ for some invertible $T \in \mathbb{C}^{n \times n}$. It is easily seen that congruent matrices form an equivalence class in which the rank is an invariant. Furthermore, congruence of hermitian matrices has the important property of preserving the “inertia”, i.e. the numbers of positive, negative, and zero eigenvalues as specified more precisely in:

Theorem 4.1 *Each equivalence class of congruent hermitian matrices in $\mathbb{C}^{n \times n}$ contains exactly one matrix with the partitioned form*

$$D = \begin{bmatrix} I_s & 0 & 0 \\ 0 & -I_{r-s} & 0 \\ 0 & 0 & 0_{n-r} \end{bmatrix}, \quad (4.1)$$

where r is the rank of all matrices in the class and s is the number of positive eigenvalues, each counted as many times as its algebraic multiplicity.

In this partitioned matrix the last row and column simply does not appear if $r = n$. Clearly, the invariant $r - s$ is just the number of negative eigenvalues (counted with algebraic multiplicities).

The integers in (4.1) constitute the *inertia* of matrices in the equivalence class of congruent hermitian matrices. Thus, if $X \in \mathbb{C}^{n \times n}$ is hermitian, and is congruent to the matrix D in (4.1), we write

$$\text{In}(X) = \{s, r - s, n - r\},$$

and

$$\text{In}_+(X) = s, \quad \text{In}_-(X) = r - s, \quad \text{In}_0(X) = n - r.$$

Clearly,

$$\text{rank}(X) = \text{In}_+(X) + \text{In}_-(X). \quad (4.2)$$

Theorem 4.1 holds also over the reals. Thus, $H_1, H_2 \in \mathbb{R}^{n \times n}$ are said to be *congruent* (over the reals) if $H_1 = T^* H_2 T$ for some invertible $T \in \mathbb{R}^{n \times n}$. Again, each equivalence class of congruent (over the reals) real symmetric matrices contains exactly one matrix in the form (4.1).

For future reference some easily verified properties of the inertia are recorded here:

Lemma 4.2 (a) *If H_1, \dots, H_p are hermitian matrices, then*

$$\text{In}_\pm(H_1 \oplus \dots \oplus H_p) = \sum_{k=1}^p \text{In}_\pm(H_k),$$

$$\text{In}_0(H_1 \oplus \cdots \oplus H_p) = \sum_{k=1}^p \text{In}_0(H_k).$$

(b) With F_m defined as in (2.5),

$$\text{In}_\pm(F_m) = \begin{cases} m/2 & \text{if } m \text{ is even} \\ (m+1)/2 & \text{if } m \text{ is odd} \end{cases}$$

Continuity of the eigenvalues (properly enumerated) of a matrix, as functions of the entries of the matrix, yields the following fact:

Theorem 4.3 *Let there be given a hermitian $X \in \mathbb{C}^{n \times n}$. Then there exists an $\epsilon > 0$ such that*

$$\text{In}_+(Y) \geq \text{In}_+(X), \quad \text{In}_-(Y) \geq \text{In}_-(X)$$

for every hermitian $Y \in \mathbb{C}^{n \times n}$ satisfying $\|Y - X\| < \epsilon$.

5 Pairs of hermitian matrices: strict equivalence

In this section we consider hermitian *matrix pencils* i.e. matrix pencils of the form $A + \lambda B$, where $A = A^*$ and $B = B^*$ are hermitian $n \times n$ matrices and λ is real. Clearly, congruent hermitian matrix pencils are strictly equivalent, but not conversely. In this and the next section the connections between congruence and strict equivalence for hermitian matrix pencils will be explored. First, a canonical form for hermitian matrix pencils under strict equivalence is obtained (Theorem 5.1) by straightforward manipulations with the Kronecker form; namely, taking adjoints and then using the uniqueness property. This is used in the next section to formulate an hermitian canonical form under congruence (Theorem 6.1).

Thus, let $A + \lambda B$ an hermitian matrix pencil, and let $A_0 + \lambda B_0$ be the Kronecker form of (3.1). Then

$$A_0 + \lambda B_0 = P(A + \lambda B)Q \tag{5.1}$$

for some invertible $P, Q \in \mathbb{C}^{n \times n}$. Taking conjugate transposes in (5.1),

$$A_0^* + \lambda B_0^* = Q^*(A + \lambda B)P^*.$$

Clearly, $A_0^* + \lambda B_0^*$ is also strictly equivalent to $A + \lambda B$. But

$$\begin{aligned} A_0^* + \lambda B_0^* &= 0_{v \times u} \oplus L_{\epsilon_1 \times (\epsilon_1 + 1)}^T \oplus \cdots \oplus L_{\epsilon_p \times (\epsilon_p + 1)}^T \\ &\quad \oplus L_{\eta_1 \times (\eta_1 + 1)} \oplus \cdots \oplus L_{\eta_q \times (\eta_q + 1)} \\ &\quad \oplus (I_{k_1} + \lambda J_{k_1}(0)^T) \oplus \cdots \oplus (I_{k_r} + \lambda J_{k_r}(0)^T) \\ &\quad \oplus (\lambda I_{\ell_1} + J_{\ell_1}(\alpha_1)^*) \oplus \cdots \oplus (\lambda I_{\ell_s} + J_{\ell_s}(\alpha_s)^*), \end{aligned}$$

which, in view of the similarity between $J_m(\alpha)^*$ and $J_m(\bar{\alpha})$ (see 2.7), is strictly equivalent to

$$\begin{aligned} & 0_{r \times u} \oplus L_{\varepsilon_1 \times (\varepsilon_1 + 1)}^T \oplus \cdots \oplus L_{\varepsilon_p \times (\varepsilon_p + 1)}^T \\ & \oplus L_{\eta_1 \times (\eta_1 + 1)} \oplus \cdots \oplus L_{\eta_q \times (\eta_q + 1)} \\ & \oplus (I_{k_1} + \lambda J_{k_1}(0)) \oplus \cdots \oplus (I_{k_r} + \lambda J_{k_r}(0)) \\ & \oplus (\lambda I_{\ell_1} + J_{\ell_1}(\bar{\alpha}_1)) \oplus \cdots \oplus (\lambda I_{\ell_s} + J_{\ell_s}(\bar{\alpha}_s)). \end{aligned} \quad (5.2)$$

Now the uniqueness of the Kronecker form implies that $r = u$; $p = q$; $\varepsilon_j = \eta_j$ for $j = 1, \dots, q$ and, also, that the number of blocks $\lambda I_{\ell_t} + J_{\ell_t}(\alpha_t)$ with α_t nonreal is even and these blocks appear in pairs: $\lambda I_{\ell_t} + J_{\ell_t}(\alpha_t)$, $\lambda I_{\ell_t} + J_{\ell_t}(\bar{\alpha}_t)$.

In order to describe some canonical forms it is necessary to define more real symmetric matrices of standard form. They include the sip matrix, F_m , of (2.5),

$$G_m = \begin{bmatrix} 0 & \cdots & \cdots & 1 & 0 \\ \vdots & & & 0 & 0 \\ \vdots & & & \vdots & \\ 1 & 0 & & \vdots & \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix} = \begin{bmatrix} F_{m-1} & 0 \\ 0 & 0 \end{bmatrix} = J_m(0)F_m \quad (5.3)$$

and

$$\tilde{G}_m = F_m G_m F_m = \begin{bmatrix} 0 & 0 \\ 0 & F_{m-1} \end{bmatrix}. \quad (5.4)$$

The goal of this analysis is an *hermitian* matrix pencil which is strictly equivalent to (5.2). To this end, symmetric expressions can be obtained from the blocks of (5.2) on multiplying by suitable nonsingular factors, i.e. by applying suitable strict equivalence transformations. Thus, for the singular part of the form (5.2), it is easily verified that

$$(L_{\varepsilon \times (\varepsilon + 1)} \oplus (F_{\varepsilon + 1} L_{\varepsilon \times (\varepsilon + 1)}^T F_{\varepsilon})) F_{2\varepsilon + 1} = \lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon} \\ 0 & 0 & 0 \\ F_{\varepsilon} & 0 & 0 \end{bmatrix} + G_{2\varepsilon + 1}.$$

Similarly, for the regular part:

$$\begin{aligned} (I_{k_t} + \lambda J_{k_t}(0)) F_{k_t} &= \lambda G_{k_t} + F_{k_t}; \\ (\lambda I_{\ell_t} + J_{\ell_t}(\alpha)) F_{\ell_t} &= (\lambda + \alpha) F_{\ell_t} + G_{\ell_t}, \end{aligned}$$

where α is real; and

$$\begin{aligned} & ((\lambda I_{\ell_t} + J_{\ell_t}(\alpha)) \oplus (\lambda I_{\ell_t} + J_{\ell_t}(\bar{\alpha}))) F_{2\ell_t} \\ &= \begin{bmatrix} 0 & (\lambda + \alpha) F_{\ell_t} \\ (\lambda + \bar{\alpha}) F_{\ell_t} & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_{\ell_t} \\ G_{\ell_t} & 0 \end{bmatrix} \end{aligned} \quad (5.5)$$

for non-real α .

Applying these transformations to the blocks in (5.2), the following result is obtained:

Theorem 5.1 *Every hermitian matrix pencil $A + \lambda B$ is strictly equivalent to a hermitian matrix pencil of the form*

$$\begin{aligned}
0_{u \times u} \oplus & \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_1} \\ 0 & 0 & 0 \\ F_{\varepsilon_1} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_1+1} \right) \oplus \cdots \oplus \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_p} \\ 0 & 0 & 0 \\ F_{\varepsilon_p} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_p+1} \right) \\
& \oplus (F_{k_1} + \lambda G_{k_1}) \oplus \cdots \oplus (F_{k_r} + \lambda G_{k_r}) \\
& \oplus ((\lambda + \alpha_1) F_{\ell_1} + G_{\ell_1}) \oplus \cdots \oplus ((\lambda + \alpha_q) F_{\ell_q} + G_{\ell_q}) \\
& \oplus \left(\begin{bmatrix} 0 & (\lambda + \beta_1) F_{m_1} \\ (\lambda + \bar{\beta}_1) F_{m_1} & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_{m_1} \\ G_{m_1} & 0 \end{bmatrix} \right) \\
& \oplus \cdots \oplus \left(\begin{bmatrix} 0 & (\lambda + \beta_s) F_{m_s} \\ (\lambda + \bar{\beta}_s) F_{m_s} & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_{m_s} \\ G_{m_s} & 0 \end{bmatrix} \right). \tag{5.6}
\end{aligned}$$

Here, $\varepsilon_1 \leq \cdots \leq \varepsilon_p$ and $k_1 \leq \cdots \leq k_r$ are positive integers, α_j are real numbers, β_j are complex nonreal numbers, and F_m, G_m are the $m \times m$ matrices given by (2.5) and (5.3).

The form (5.5) is uniquely determined by $A + \lambda B$ up to a combination of the following permutations: a permutation of the blocks

$$\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_j} \\ 0 & 0 & 0 \\ F_{\varepsilon_j} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_j+1}, \quad j = 1, \dots, p;$$

a permutation of the blocks

$$(\lambda + \alpha_j) F_{\ell_j} + G_{\ell_j}, \quad j = 1, \dots, q;$$

a permutation of the blocks

$$F_{k_j} + \lambda G_{k_j}, \quad j = 1, \dots, r;$$

and a permutation of the blocks

$$\begin{bmatrix} 0 & (\lambda + \beta_j) F_{m_j} \\ (\lambda + \bar{\beta}_j) F_{m_j} & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_{m_j} \\ G_{m_j} & 0 \end{bmatrix} \quad (j = 1, \dots, s)$$

with possible replacement of β_j by $\bar{\beta}_j$ within each such block.

An alternate form can be obtained instead of (5.5) in which the $m \times m$ symmetric matrix \tilde{G}_m of (5.4) is used in place of G_m , and $\begin{bmatrix} 0 & \tilde{G}_m \\ \tilde{G}_m & 0 \end{bmatrix}$ is used in place of $\begin{bmatrix} 0 & G_m \\ G_m & 0 \end{bmatrix}$. The verification of the alternate form is left to the reader (see the remarks after Theorem 6.1 in the next section for more detail).

6 Pairs of hermitian matrices: canonical form under congruence

Congruence transformations are, of course, more restrictive than those of strict equivalence. So it is to be expected that the canonical forms that can be attained (for square pencils) will be more complex. It is surprising that, in fact, these canonical forms are more complex in only one respect: the introduction of a so-called “sign-characteristic” associated with the real (and infinite) eigenvalues. The block structures achieved using congruence are just those of Theorem 5.1. In the following statement new parameters δ and η are associated with the eigenvalue at infinity and the finite real eigenvalues, respectively.

Theorem 6.1 *Every hermitian matrix pencil $A + \lambda B$ is congruent to a hermitian matrix pencil of the form*

$$\begin{aligned}
0_{u \times u} \oplus & \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_1} \\ 0 & 0 & 0 \\ F_{\varepsilon_1} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_1+1} \right) \oplus \cdots \oplus \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_p} \\ 0 & 0 & 0 \\ F_{\varepsilon_p} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_p+1} \right) \\
& \oplus \delta_1 (F_{k_1} + \lambda G_{k_1}) \oplus \cdots \oplus \delta_r (F_{k_r} + \lambda G_{k_r}) \\
& \oplus \eta_1 ((\lambda + \alpha_1) F_{l_1} + G_{l_1}) \oplus \cdots \oplus \eta_q ((\lambda + \alpha_q) F_{l_q} + G_{l_q}) \\
& \oplus \left(\begin{bmatrix} 0 & (\lambda + \beta_1) F_{m_1} \\ (\lambda + \bar{\beta}_1) F_{m_1} & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_{m_1} \\ G_{m_1} & 0 \end{bmatrix} \right) \\
& \oplus \cdots \oplus \left(\begin{bmatrix} 0 & (\lambda + \beta_s) F_{m_s} \\ (\lambda + \bar{\beta}_s) F_{m_s} & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_{m_s} \\ G_{m_s} & 0 \end{bmatrix} \right). \tag{6.1}
\end{aligned}$$

Here, $\varepsilon_1 \leq \cdots \leq \varepsilon_p$ and $k_1 \leq \cdots \leq k_r$ are positive integers, α_j are real numbers, β_j are complex nonreal numbers, $\delta_1, \dots, \delta_r, \eta_1, \dots, \eta_q$ are signs, each equal to $+1$ or -1 , and F_m is the sip matrix (5.3).

The form (6.1) is uniquely determined by $A + \lambda B$ up to a combination of permutations of the following types: a permutation of the blocks

$$\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_j} \\ 0 & 0 & 0 \\ F_{\varepsilon_j} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_j+1}, \quad j = 1, 2, \dots, p;$$

a permutation of the blocks

$$\delta_j (F_{k_j} + \lambda G_{k_j}), \quad j = 1, \dots, r;$$

a permutation of the blocks

$$\eta_j ((\lambda + \alpha_j) F_{l_j} + G_{l_j}), \quad j = 1, \dots, q;$$

and a permutation of the blocks

$$\begin{bmatrix} 0 & (\lambda + \beta_j)F_{m_j} \\ (\lambda + \bar{\beta}_j)F_{m_j} & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_{m_j} \\ G_{m_j} & 0 \end{bmatrix} \quad (j = 1, \dots, s), \quad (6.2)$$

with possible replacement of β_j by $\bar{\beta}_j$ within each such block.

This theorem has a long history including a series of papers (that appeared more or less simultaneously) by Trott [53], Turnbull [54], Ingraham and Wegner [22], [58], [21], and Williamson [59]. It was rediscovered later and generalized in many ways. We mention here just two more early works by Williamson [61] and Dieudonne [8]. An extensive bibliography is given in [52].

The sequence of signs $\{\delta_1, \dots, \delta_r; \eta_1, \dots, \eta_q\}$ appearing here is called the *sign characteristic* of the hermitian matrix pencil $A + \lambda B$.

An alternate version of Theorem 6.1 is obtained, just as for Theorem 5.1, by using \tilde{G}_α in (6.1) in the place of G_α , and

$$\begin{bmatrix} 0 & \tilde{G}_\alpha \\ \tilde{G}_\alpha & 0 \end{bmatrix} \text{ in place of } \begin{bmatrix} 0 & G_\alpha \\ G_\alpha & 0 \end{bmatrix} \quad (\beta \neq \bar{\beta}).$$

Again, the full verification of the alternative version is left to the reader, and we only indicate the following relevant easily verifiable equalities:

$$F_{2\varepsilon+1} \left(\lambda \begin{bmatrix} 0 & 0 & F_\varepsilon \\ 0 & 0 & 0 \\ F_\varepsilon & 0 & 0 \end{bmatrix} + G_{2\varepsilon+1} \right) = \left(\lambda \begin{bmatrix} 0 & 0 & F_\varepsilon \\ 0 & 0 & 0 \\ F_\varepsilon & 0 & 0 \end{bmatrix} + \tilde{G}_{2\varepsilon+1} \right) F_{2\varepsilon+1};$$

$$F_k (F_k + \lambda G_k) = (F_k + \lambda \tilde{G}_k) F_k;$$

$$F_{2m} \left(\begin{bmatrix} 0 & (\lambda + \beta)F_m \\ (\lambda + \bar{\beta})F_m & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_m \\ G_m & 0 \end{bmatrix} \right)$$

$$= \left(\begin{bmatrix} 0 & (\lambda + \bar{\beta})F_m \\ (\lambda + \beta)F_m & 0 \end{bmatrix} + \begin{bmatrix} 0 & \tilde{G}_m \\ \tilde{G}_m & 0 \end{bmatrix} \right) F_{2m}, \quad \beta \neq \bar{\beta}.$$

We also indicate an explicit congruence of the block (6.2) to the analogous block in which β_j is replaced by $\bar{\beta}_j$. Denoting the block (6.2) by $K_{2m_j}(\beta_j)$, we have:

$$\begin{bmatrix} 0 & I_{m_j} \\ I_{m_j} & 0 \end{bmatrix} K_{2m_j}(\beta_j) \begin{bmatrix} 0 & I_{m_j} \\ I_{m_j} & 0 \end{bmatrix} = K_{2m_j}(\bar{\beta}_j).$$

The complete proof of Theorem 6.1 is rather long and will be given in the next two sections. Here we deduce some immediate corollaries by comparing Theorems 6.1 and 5.1. It will be confirmed in Section 9 that similar results hold for real symmetric pencils under real congruence.

Corollary 6.2 *The strict equivalence class of a hermitian matrix pencil $A + \lambda B$, where $A = A^*, B = B^* \in \mathbb{C}^{n \times n}$, contains only a finite number of congruence classes of $A + \lambda B$. In fact, the number m of these congruence classes is bounded above by a constant $C(n)$ that depends on n only, namely,*

$$C(n) = \max [(s_1 + 1) \dots (s_p + 1)], \quad (6.3)$$

where the maximum is taken over all possible distinct positive integers m_1, \dots, m_p and all possible positive integers s_1, \dots, s_p such that

$$s_1 m_1 + \dots + s_p m_p = n. \quad (6.4)$$

Moreover, the bound (6.3) for m is achieved.

For example, $C(5) = 8$. If equality (6.4) is considered without the requirement that m_1, \dots, m_p be distinct, then it is easy to see that the maximum of $(s_1 + 1) \dots (s_p + 1)$ is attained when all the m_p 's and all the s_p 's are equal to 1. Thus, $C(n) \leq 2^n$.

Corollary 6.3 *The following properties are equivalent for a hermitian matrix pencil $A + \lambda B$:*

(1) *Every hermitian matrix pencil which is strictly equivalent to $A + \lambda B$ is also congruent to $A + \lambda B$;*

(2) *The Jordan parts of $A + \lambda B$ and of $B + \lambda A$ have no real eigenvalues.*

(3)

$$\text{rank}(A + \lambda B) = \text{rank } B \quad \text{for all real } \lambda.$$

(4)

$$\text{In}(A + \lambda B) = \text{In}(B) \quad \text{for all real } \lambda.$$

Proof. In view of Theorems 6.1 and 4.1, we may use the form (6.1) in place of $A + \lambda B$. It follows from (6.1) that (1) holds if and only if the parts

$$\delta_1 (F_{k_1} + \lambda G_{k_1}) \oplus \dots \oplus \delta_r (F_{k_r} + \lambda G_{k_r})$$

and

$$\eta_1 ((\lambda + \alpha_1) F_{\ell_1} + G_{\ell_1}) \oplus \dots \oplus \eta_q ((\lambda + \alpha_q) F_{\ell_q} + G_{\ell_q})$$

are absent in (6.1). By inspection, this is equivalent to property (3), and by definition of the Jordan parts of $A + \lambda B$ and of $B + \lambda A$, this is also equivalent to (2).

It remains to prove that (3) is equivalent to (4). Clearly, (4) implies (3). The reverse implication follows easily by using (4.2) and Theorem 4.3. □

If $X \in \mathbb{C}^{n \times n}$, then by applying Theorem 6.1 to the pencil $A + \lambda B$, where A and B are hermitian matrices such that $X = A + iB$, the following canonical form under congruence is obtained:

Corollary 6.4 *Every matrix $X \in \mathbb{C}^{n \times n}$ is congruent to a matrix of the form (6.1), where λ is replaced by i . The form (6.1) (with λ replaced by i) is unique up to a combination of permutations as described in Theorem 6.1.*

7 Proof of Theorem 6.1: existence

The existence of the form (6.1) will be proved here by taking advantage of Theorem 5.1. The uniqueness property will be established in the next section.

We state a well-known result on solutions of linear matrix equations that will be used in the proof. The result applies in both the real and the complex cases.

Theorem 7.1 *Let A and B be matrices of sizes $m \times m$ and $n \times n$, respectively. Then the matrix equation*

$$AX - XB = 0 \tag{7.1}$$

has only the trivial solution $X = 0_{m \times n}$ if and only if $\sigma(A) \cap \sigma(B) = \emptyset$.

For a proof see, for example, [31, Chapter 12], where the complex case is considered. However, if A and B are real matrices then, since the equation (7.1) is linear, it follows that (7.1) has only the trivial solution over the reals if and only if it has only the trivial solution over the complexes.

Let $A + \lambda B$ be a hermitian matrix pencil. By Theorem 5.1, there exist invertible matrices P and Q such that

$$A + \lambda B = P(A_0 + \lambda B_0)Q, \tag{7.2}$$

where $A_0 + \lambda B_0$ is the canonical form of (5.6). Replacing $A + \lambda B$ by the congruent pencil $(Q^*)^{-1}(A + \lambda B)Q^{-1}$, it may be assumed without loss of generality that $Q = I$. Then the equalities $A = A^*$, $B = B^*$ imply

$$PA_0 = A_0P^*, \quad PB_0 = B_0P^*. \tag{7.3}$$

Therefore, for every polynomial $f(\lambda)$ with real coefficients we also have

$$f(P)A_0 = A_0(f(P))^*, \quad f(P)B_0 = B_0(f(P))^*. \tag{7.4}$$

The argument now proceeds in several steps depending on the nature of the spectrum of P . An important role is played by a critical lemma attributed by Thompson to Hua [20], and which appears below as Lemma 7.2.

Case 1 *P has only one distinct eigenvalue.*

It will be shown that this eigenvalue is necessarily real (unless both A and B are equal to zero, in which case Theorem 6.1 is trivial). Indeed, assuming $A_0 \neq 0$, we have

$$A_0 = U^* \begin{bmatrix} 0 & 0 \\ 0 & Y \end{bmatrix} U$$

for some unitary U and invertible hermitian Y . Letting $P_0 := UPU^* = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$, where P_{22} has the same size as Y , the equality $PA_0 = A_0P^*$ becomes

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & Y \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & Y \end{bmatrix} \begin{bmatrix} P_{11}^* & P_{21}^* \\ P_{12}^* & P_{22}^* \end{bmatrix}.$$

This equality implies $P_{12} = 0$ and $P_{22}Y = YP_{22}^*$. In particular, the eigenvalues of P_{22} are also eigenvalues of P , and hence P_{22} has only one distinct eigenvalue. But the equality $P_{22}Y = YP_{22}^*$ means that P_{22}^* is similar to P_{22} . It follows from the Jordan form for P_{22} that the nonreal eigenvalues of P_{22}^* must arise in complex conjugate pairs. Thus, as P_{22} has precisely one distinct eigenvalue, it must be real - and the same is true for P .

Assume first that the eigenvalue of P is positive. Let $(\lambda - \gamma)^n$, $\gamma > 0$, be the characteristic polynomial of P^{-1} . By Lemma 7.2 (below) there exists a polynomial $f_n(\lambda)$ with real coefficients such that $\lambda = (f_n(\lambda))^2 \pmod{(\lambda - \gamma)^n}$. Then $P^{-1} = (f_n(P^{-1}))^2$, and since P^{-1} itself is a polynomial of P with real coefficients, $P^{-1} = (f(P))^2$ for some polynomial $f(\lambda)$ with real coefficients. Let $R = f(P)$, and use (7.4) to obtain

$$\begin{aligned} R(A + \lambda B)R^* &= RP(A_0 + \lambda B_0)R^* = RP(A_0 + \lambda B_0)(f(P))^* \\ &= f(P)Pf(P)(A_0 + \lambda B_0) = A_0 + \lambda B_0, \end{aligned} \quad (7.5)$$

and the existence of (6.1) follows.

Now assume that the eigenvalue of P is negative. Then, arguing as in the preceding paragraph, it is found that $(-P)^{-1} = (f(-P))^2$ for some polynomial $f(\lambda)$ with real coefficients. As in (7.5) it follows that

$$R(A + \lambda B)R^* = -(A_0 + \lambda B_0),$$

where $R = f(-P)$. Thus, to complete the proof of existence of the form (6.1) in Case 1, it remains to show that each of the blocks

$$Z_1 = \lambda \begin{bmatrix} 0 & 0 & F_\varepsilon \\ 0 & 0 & 0 \\ F_\varepsilon & 0 & 0 \end{bmatrix} + G_{2\varepsilon+1} \quad (7.6)$$

and

$$Z_2 = \begin{bmatrix} 0 & (\lambda + \beta)F_m \\ (\lambda + \bar{\beta})F_m & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_m \\ G_m & 0 \end{bmatrix}, \quad \beta \in \mathbb{C} \setminus \mathbb{R}, \quad (7.7)$$

is congruent to its negative. This is easily seen because, for the block (7.7) we have

$$\begin{bmatrix} I_m & 0 \\ 0 & -I_m \end{bmatrix} Z_2 \begin{bmatrix} I_m & 0 \\ 0 & -I_m \end{bmatrix} = -Z_2,$$

and for the block (7.6) we have

$$\begin{bmatrix} I_\varepsilon & 0 \\ 0 & -I_{\varepsilon+1} \end{bmatrix} Z_1 \begin{bmatrix} I_\varepsilon & 0 \\ 0 & -I_{\varepsilon+1} \end{bmatrix} = -Z_1. \quad (7.8)$$

Case 2 *P has exactly two distinct eigenvalues, and they are complex conjugates of each other.*

Let $\alpha \neq \bar{\alpha}$ be the eigenvalues of P^{-1} . If w is the maximum of the algebraic multiplicities of α and $\bar{\alpha}$, then P^{-1} satisfies $g(P^{-1}) = 0$, where $g(\lambda) = ((\lambda - \alpha)(\lambda - \bar{\alpha}))^w$. Using Lemma 7.2, we find a polynomial $g_w(\lambda)$ with real coefficients such that $\lambda = (g_w(\lambda))^2 \pmod{g(\lambda)}$. Then $P^{-1} = (g_w(P^{-1}))^2$. Since P is itself a matrix root of a polynomial with real coefficients, namely, $\widehat{g}(P) = 0$, where

$$\widehat{g}(\lambda) = ((\lambda - \alpha^{-1})(\lambda - (\bar{\alpha})^{-1}))^w,$$

it follows that P^{-1} is also a polynomial of P with real coefficients. Now argue as in Case 1 (when the eigenvalue of P is positive).

Case 3 *All other possibilities (not covered in Cases 1 and 2).*

In this case, using the complex Jordan form of Theorem 2.1, P can be written in the form

$$P = S \begin{bmatrix} P_1 & & & 0 \\ & P_2 & & \\ & & \ddots & \\ 0 & & & P_r \end{bmatrix} S^{-1}, \quad (7.9)$$

where S is invertible and P_1, \dots, P_r are matrices of sizes $n_1 \times n_1, \dots, n_r \times n_r$, respectively, such that

$$\lambda \in \sigma(P_i) \Rightarrow \bar{\lambda} \notin \sigma(P_j) \quad \text{for } j \neq i, \quad (7.10)$$

(i.e. any conjugate pair is confined to just one block, P_j). We also have $r \geq 2$ (the situations when $r = 1$ are covered in cases 1 and 2). Substituting (7.9) in the equality $A + \lambda B = P(A_0 + \lambda B_0)$, we obtain

$$(P_1^{-1} \oplus \dots \oplus P_r^{-1})(\tilde{A} + \lambda \tilde{B}) = \tilde{A}_0 + \lambda \tilde{B}_0, \quad (7.11)$$

where

$$\tilde{A} = S^{-1}A(S^*)^{-1}, \quad \tilde{B} = S^{-1}B(S^*)^{-1}, \quad \tilde{A}_0 = S^{-1}A_0(S^*)^{-1}, \quad \tilde{B}_0 = S^{-1}B_0(S^*)^{-1}.$$

Partition the matrix \tilde{A} :

$$\tilde{A} = [M_{ij}]_{i,j=1}^r,$$

where M_{ij} is of the size $n_i \times n_j$. Since \tilde{A} and \tilde{A}_0 are hermitian, (7.11) implies

$$P_i^{-1}M_{ij} = M_{ij}(P_j^*)^{-1}.$$

In view of (7.10),

$$\sigma(P_i^{-1}) \cap \sigma((P_j^*)^{-1}) = \emptyset \quad \text{for } i \neq j.$$

Now by Theorem 7.1 we have $M_{ij} = 0$ ($i \neq j$). In other words,

$$\tilde{A} = M_{11} \oplus \dots \oplus M_{rr}.$$

Similarly, $\tilde{B} = N_{11} \oplus \dots \oplus N_{rr}$, where N_{ii} is of size $n_i \times n_i$.

Now induction is used on the size of the pencil $A + \lambda B$ to complete the proof that every hermitian matrix pencil is congruent to a pencil of the form (6.1); the basis of induction, when A and B are scalars, is trivially verified. Indeed, by the induction hypothesis, each pencil $M_{ii} + \lambda N_{ii}$ is congruent to a pencil of the form (6.1), and therefore the same is true for

$$\tilde{A} + \lambda \tilde{B} = (M_{11} + \lambda N_{11}) \oplus \dots \oplus (M_{rr} + \lambda N_{rr}).$$

This completes the proof of existence of the form (6.1).

Lemma 7.2 *Let $p(\lambda)$ be a real scalar polynomial having one of the forms $p(\lambda) = \lambda - a$ with $a > 0$, or $p(\lambda) = (\lambda - a)(\lambda - \bar{a})$ with $a \in \mathbb{C} \setminus \mathbb{R}$. Then for every positive integer m there exists a scalar polynomial $f_m(\lambda)$ with real coefficients such that $\lambda \equiv (f_m(\lambda))^2 \pmod{(p(\lambda))^m}$.*

Proof. For $m = 1$, we let $f_1(\lambda) = \sqrt{a}$ if $p(\lambda) = \lambda - a$, and

$$f_1(\lambda) = \{|a| + \lambda\} \{2(|a| + \operatorname{Re} a)\}^{-1/2}$$

if $p(\lambda) = (\lambda - a)(\lambda - \bar{a})$. Continue by induction on m . If $f_m(\lambda)$ has already been found, then

$$\lambda = f_m(\lambda)^2 + t(\lambda)(p(\lambda))^m \tag{7.12}$$

for some real polynomial $t(\lambda)$. In particular, (7.12) implies that $f_m(\lambda)$ and $p(\lambda)$ are relatively prime. Therefore, there exist polynomials $g(\lambda)$ and $h(\lambda)$ with real coefficients such that

$$2g(\lambda)f_m(\lambda) + h(\lambda)p(\lambda) = t(\lambda).$$

Now set

$$f_{m+1}(\lambda) = f_m(\lambda) + g(\lambda)(p(\lambda))^m.$$

□

8 Pairs of hermitian matrices: uniqueness of the canonical form

We now prove the uniqueness of the form (6.1) under congruence. A lemma is to be proved first which shows that congruence of two pencils implies congruence of corresponding pairs of canonical blocks in (6.1).

Recall the definitions of singular and regular pencils made in Section 3. Thus, a hermitian matrix pencil $A + \lambda B$ is *singular* if $\det(A + \lambda B) \equiv 0$ and *regular* otherwise. A *singular* hermitian matrix pencil $A + \lambda B$ is said to have *finite indices* if it is congruent to a pencil of the form

$$\bigoplus_{j=1}^p \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_j} \\ 0 & 0_1 & 0 \\ F_{\varepsilon_j} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_j+1} \right),$$

so that only the first block-row of equation (6.1) appears.

Lemma 8.1 *Let two hermitian matrix pencils be given in the form*

$$A_1 + \lambda B_1 = 0_{u \times u} \oplus (A_{10} + \lambda B_{10}) \oplus (A_{11} + \lambda B_{11}) \oplus \bigoplus_{j=2}^m (A_{1j} + \lambda B_{1j}) \quad (8.1)$$

and

$$A_2 + \lambda B_2 = 0_{u \times u} \oplus (A_{20} + \lambda B_{20}) \oplus (A_{21} + \lambda B_{21}) \oplus \bigoplus_{j=2}^m (A_{2j} + \lambda B_{2j}), \quad (8.2)$$

where $A_{1j} + \lambda B_{1j}$ and $A_{2j} + \lambda B_{2j}$ are of the same size $n_j \times n_j$ ($j = 0, \dots, m$), and in addition have the following properties: $A_{k0} + \lambda B_{k0}$ ($k = 1, 2$) are singular hermitian matrix pencils with finite indices, the determinants of $A_{k1} + \lambda B_{k1}$ are non-zero constants, and the determinants of $A_{kj} + \lambda B_{kj}$ ($k = 1, 2; j = 2, \dots, m$) are nonconstant polynomials of λ satisfying the following properties:

- (i) the degree of $\det(A_{kj} + \lambda B_{kj})$ is equal to n_j ($j = 2, \dots, m$);
- (ii) $\det(A_{1j} + \lambda B_{1j}) = c_j \det(A_{2j} + \lambda B_{2j})$, $j = 2, \dots, m$, where c_j are nonzero constants.
- (iii) For $k = 1$ and $k = 2$, $\det(A_{kj_1} + \lambda B_{kj_1})$ and $\det(A_{kj_2} + \lambda B_{kj_2})$ are relatively prime for every pair of indices $2 \leq j_1 < j_2 \leq m$.

Assume also that $A_1 + \lambda B_1$ and $A_2 + \lambda B_2$ are congruent. Then for each $j = 0, 1, \dots, m$, the hermitian matrix pencils $A_{1j} + \lambda B_{1j}$ and $A_{2j} + \lambda B_{2j}$ are congruent.

Proof. Using the existence part of Theorem 6.1, we may partition $A_k + \lambda B_k$ differently (replacing, if necessary, $A_k + \lambda B_k$ by a congruent hermitian matrix pencil)

$$A_k + \lambda B_k = 0_{u \times u} \oplus \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_1} \\ 0 & 0 & 0 \\ F_{\varepsilon_1} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_1+1} \right) \oplus \cdots \oplus \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_p} \\ 0 & 0 & 0 \\ F_{\varepsilon_p} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_p+1} \right) \oplus (M_{k1} + \lambda N_k) \oplus (M_{k2} + \lambda N_{k2}), \quad (8.3)$$

where M_{k1} and N_{k2} are invertible; thus, $M_{k1} + \lambda N_{k1} = A_{k1} + \lambda B_{k1}$ and

$$M_{k2} + \lambda N_{k2} = (A_{k2} + \lambda B_{k2}) \oplus (A_{k3} + \lambda B_{k3}) \oplus \cdots \oplus (A_{km} + \lambda B_{km})$$

The uniqueness part of Theorem 6.1 guarantees that the indices u and $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ are indeed the same for $k = 1$ and for $k = 2$. In particular, $A_{10} + \lambda B_{10}$ and $A_{20} + \lambda B_{20}$ are congruent.

Let

$$T(A_1 + \lambda B_1)T^* = A_2 + \lambda B_2 \quad (8.4)$$

for some invertible T . Rewrite (8.4) in the form

$$T(A_1 + \lambda B_1) = (A_2 + \lambda B_2)S, \quad (8.5)$$

where $S = (T^*)^{-1}$, and partition T and S conformably with (8.3):

$$T = [T^{(ij)}]_{i,j=0}^{p+2}, \quad S = [S^{(ij)}]_{i,j=0}^{p+2},$$

where $T^{(00)}$ and $S^{(00)}$ are of the size $u \times u$, $T^{(ii)}$ and $S^{(ii)}$ are of the size $(2\varepsilon_i+1) \times (2\varepsilon_i+1)$ for $i = 1, \dots, p$, and $T^{(p+j,p+j)}$ and $S^{(p+j,p+j)}$ have the same size as $M_{kj} + \lambda N_{kj}$ ($j = 1, 2$). Consider the blocks $T^{(p+j,m)}$, $m = 1, \dots, p$; $j = 1, 2$. Form the partitions:

$$\begin{aligned} T^{(p+j,m)} &= [V^{(p+j,m)} U^{(p+j,m)}], \\ S^{(p+j,m)} &= [Y^{(p+j,m)} Z^{(p+j,m)}], \end{aligned}$$

where $V^{(p+j,m)}$ and $Y^{(p+j,m)}$ have ε_m columns. Equating the $(p+j, m)$ blocks in both sides of (8.5), we obtain

$$T^{(p+j,m)} \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_m} \\ 0 & 0 & 0 \\ F_{\varepsilon_m} & 0 & 0 \end{bmatrix} + \begin{bmatrix} F_{2\varepsilon_m} & 0 \\ 0 & 0 \end{bmatrix} \right) = (M_j + \lambda N_{2j})S^{(p+j,m)}.$$

Therefore

$$V^{(p+j,m)}([0 \quad \lambda F_{\varepsilon_m}] + [F_{\varepsilon_m} \quad 0]) = (M_{2j} + \lambda N_{2j})Z^{(p+j,m)},$$

or

$$V^{(p+j,m)} [F_{\varepsilon_m} \quad 0] = M_{2j} Z^{(p+j,m)}, \quad V^{(p+j,m)} [0 \quad F_{\varepsilon_m}] = N_{2j} Z^{(p+j,m)}. \quad (8.6)$$

If $j = 1$, then M_{2j} is invertible, and (8.6) gives

$$V^{(p+j,m)} [0 \quad F_{\varepsilon_m}] = N_{2j} M_{2j}^{-1} V^{(p+j,m)} [F_{\varepsilon_m} \quad 0]. \quad (8.7)$$

Comparison of columns in both sides of (8.7), starting with the first column, leads to the equality $V^{(p+j,m)} = 0$. If $j = 2$, then N_{2j} is invertible, and we obtain $V^{(p+j,m)} = 0$ in a similar fashion.

Denote

$$K_{\varepsilon_m} = [F_{\varepsilon_m} \quad 0] + [0 \quad \lambda F_{\varepsilon_m}].$$

Now, since $V^{(p+j,m)} = 0$, ($j = 1, 2$; $m = 1, \dots, p$), it is found that

$$\begin{aligned} & (M_{21} + \lambda N_{21}) \oplus (M_{22} + \lambda N_{22}) \\ = & \begin{bmatrix} T^{(p+1,0)} & T^{(p+1,1)} & \dots & T^{(p+1,p+2)} \\ T^{(p+2,0)} & T^{(p+2,1)} & \dots & T^{(p+2,p+2)} \end{bmatrix} (A_1 + \lambda B_1) \begin{bmatrix} T^{(p+1,0)*} & T^{(p+2,0)*} \\ T^{(p+1,1)*} & T^{(p+2,1)*} \\ \vdots & \\ T^{(p+1,p+2)*} & T^{(p+2,p+2)*} \end{bmatrix} \\ = & \sum_{m=1}^p \begin{bmatrix} 0 & U^{(p+1,m)} \\ 0 & U^{(p+2,m)} \end{bmatrix} \begin{bmatrix} 0 & K_{\varepsilon_m} \\ K_{\varepsilon_m}^* & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ U^{(p+1,m)*} & U^{(p+2,m)*} \end{bmatrix} \\ & + \tilde{T} ((M_{11} + \lambda N_{11}) \oplus (M_{12} + \lambda N_{12})) \tilde{T}^*, \end{aligned} \quad (8.8)$$

where

$$\tilde{T} = \begin{bmatrix} T^{(p+1,p+1)} & T^{(p+1,p+2)} \\ T^{(p+2,p+1)} & T^{(p+2,p+2)} \end{bmatrix}.$$

The first sum in (8.8) is obviously zero, and since $\det((M_{21} + \lambda N_{21}) \oplus (M_{22} + \lambda N_{22}))$ is not identically zero, it follows that \tilde{T} is non-singular. In other words, the hermitian matrix pencils

$$(M_{11} + \lambda N_{11}) \oplus (M_{12} + \lambda N_{12}) \quad \text{and} \quad (M_{21} + \lambda N_{21}) \oplus (M_{22} + \lambda N_{22})$$

are congruent.

We now return to the partitions (8.1) and (8.2). In view of the arguments made so far, it may be assumed that the blocks $0_{u \times u}$ and $A_{k0} + \lambda B_{k0}$ do not appear, i.e.

$$A_k + \lambda B_k = \bigoplus_{q=1}^m (A_{kq} + \lambda B_{kq}), \quad k = 1, 2. \quad (8.9)$$

Write

$$\tilde{T}(A_1 + \lambda B_1) = (A_2 + \lambda B_2) \tilde{S} \quad (8.10)$$

for some invertible matrix \tilde{T} , where $\tilde{S} = \tilde{T}^{*-1}$, and partition \tilde{T} and \tilde{S} conformably with (8.9):

$$\tilde{T} = \left[\tilde{T}_{ij} \right]_{i,j=1}^m, \quad \tilde{S} = \left[\tilde{S}_{ij} \right]_{i,j=1}^m.$$

Now (8.9) gives

$$\tilde{T}_{ij}A_{1j} = A_{2i}\tilde{S}_{ij}; \quad \tilde{T}_{ij}B_{1j} = B_{2i}\tilde{S}_{ij}; \quad i, j = 1, \dots, q. \quad (8.11)$$

If $i \neq 1$, then B_{2i} is invertible, and (8.11) implies

$$\tilde{T}_{ij}A_{1j} = A_{2i}B_{2i}^{-1}\tilde{T}_{1j}B_{1j}. \quad (8.12)$$

If, in addition, $j \neq 1$, then B_{1j} is also invertible, and we obtain

$$\tilde{T}_{ij}A_{1j}B_{1j}^{-1} = A_{2i}B_{2i}^{-1}\tilde{T}_{ij}. \quad (8.13)$$

But

$$\det(\lambda I - A_{1j}B_{1j}^{-1}) = \det(B_{1j}^{-1}) \det(\lambda B_{1j} - A_{1j}) = \pm \det(B_{1j}^{-1}) \det(\mu B_{1j} + A_{1j}),$$

where $\mu = -\lambda$, and similarly for $\det(\lambda I - A_{2i}B_{2i}^{-1})$. In view of the conditions (ii) and (iii), $A_{1j}B_{1j}^{-1}$ and $A_{2i}B_{2i}^{-1}$ have no common eigenvalues if $i, j \neq 1$ and $i \neq j$. Now in view of (8.13) and Theorem 7.1 we obtain $\tilde{T}_{ij} = 0$ for $i, j \neq 1$ and $i \neq j$. Let now $i \neq 1$ and $j = 1$. Then A_{11} is invertible, and (8.12) implies

$$\tilde{T}_{i1} = A_{2i}B_{2i}^{-1}\tilde{T}_{i1}B_{11}A_{11}^{-1}. \quad (8.14)$$

Now

$$\det(I + \lambda B_{11}A_{11}^{-1}) = \det(A_{11})^{-1} \det(A_{11} + \lambda B_{11})$$

is a constant, therefore $B_{11}A_{11}^{-1}$ is nilpotent: $(B_{11}A_{11}^{-1})^r = 0$ for some positive integer r . Iterating the equality (8.14), we obtain

$$\tilde{T}_{i1} = (A_{2i}B_{2i}^{-1})^r \tilde{T}_{i1} (B_{11}A_{11}^{-1})^r = 0 \quad (i \neq 1).$$

Similarly, it can be shown that $\tilde{T}_{1i} = 0$ for $i \neq 1$. Consequently, \tilde{T} is block diagonal. This completes the proof of Lemma 8.1. □

Return now to the proof of uniqueness in Theorem 6.1. Let $A + \lambda B$ be a hermitian matrix pencil which is congruent to two forms (6.1). The uniqueness part of Theorem 5.1 guarantees that apart from permutations of blocks, these two forms can possibly differ only in the signs δ_j and η_k . Lemma 8.1 allows us to reduce the proof to cases when either only blocks of the form

$$\bigoplus_{j=1}^r \delta_j (F_{k_j} + G_{k_j}) \quad (8.15)$$

are present, or only the blocks of the form

$$\bigoplus_{j=1}^q \eta_j ((\lambda + \alpha)F_{\ell_j} + G_{\ell_j}), \quad \alpha \in \mathbb{R} \quad (8.16)$$

are present. Consider these two cases separately.

Start with the form (8.16). In this case it is assumed that $A + \lambda B$ is congruent to (8.16), as well as to (8.16) with possibly different signs $\tilde{\eta}_j$. It is to be proved that, in fact, the form

$$\bigoplus_{j=1}^q \tilde{\eta}_j ((\lambda + \alpha)F_{\ell_j} + G_{\ell_j}) \quad (8.17)$$

is obtained from (8.16) after a permutation of blocks. Write

$$T \left(\bigoplus_{j=1}^q \eta_j ((\lambda + \alpha)F_{\ell_j} + G_{\ell_j}) \right) = \left(\bigoplus_{j=1}^q \tilde{\eta}_j ((\lambda + \alpha)F_{\ell_j} + G_{\ell_j}) \right) S,$$

where T is invertible, $S = (T^*)^{-1}$, and partition

$$T = [T_{ij}]_{i,j=1}^q, \quad S = [S_{ij}]_{i,j=1}^q,$$

where T_{ij} and S_{ij} are $\ell_i \times \ell_j$. Then

$$\begin{aligned} T_{ij} (\eta_j F_{\ell_j}) &= (\tilde{\eta}_i F_{\ell_i}) S_{ij}, \\ T_{ij} (\eta_j (\alpha F_{\ell_j} + G_{\ell_j})) &= (\eta_j (\alpha F_{\ell_i} + G_{\ell_j})) S_{ij}, \end{aligned}$$

and therefore

$$T_{ij} V_j = V_i T_{ij}, \quad (8.18)$$

where

$$V_i = (\alpha F_{\ell_i} + G_{\ell_i}) F_{\ell_i}^{-1} = J_{\ell_i}(\alpha).$$

Closer inspection of (8.18) shows that T_{ij} has the form

$$T_{ij} = \begin{bmatrix} 0 & \tilde{T}_{ij} \end{bmatrix} \quad (\text{if } \ell_i \leq \ell_j) \quad (8.19)$$

or

$$T_{ij} = \begin{bmatrix} \tilde{T}_{ij} \\ 0 \end{bmatrix} \quad (\text{if } \ell_i \geq \ell_j), \quad (8.20)$$

where \tilde{T}_{ij} is an upper triangular matrix of size $\min(\ell_i, \ell_j) \times \min(\ell_i, \ell_j)$ such that along each diagonal of \tilde{T}_{ij} which is parallel to the main diagonal all entries are equal. Permuting blocks in (8.16) if necessary, it can be assumed that the sizes ℓ_j are arranged in nondecreasing order. Let $u < v$ be indices such that

$$\ell_i < \ell_{u+1} = \ell_{u+2} = \cdots = \ell_v < \ell_j \quad \text{for } i \leq u, j > v.$$

Now for $u < i, k \leq v$ we have (here δ_{ik} is the Kronecker symbol: $\delta_{ik} = 1$ if $i = k$ and $\delta_{ik} = 0$ if $i \neq k$):

$$\begin{aligned} \delta_{ik} \tilde{\eta}_i F_{\ell_i} &= \sum_{j=1}^q T_{ij} (\eta_j F_{\ell_j}) T_{kj}^* = \sum_{j=1}^u T_{ij} (\eta_j F_{\ell_j}) T_{kj}^* + \\ &\quad \sum_{j=u+1}^v T_{ij} (\eta_j F_{\ell_j}) T_{kj}^* + \sum_{j=v+1}^q T_{ij} (\eta_j F_{\ell_j}) T_{kj}^*. \end{aligned} \quad (8.21)$$

In view of (8.20) the lower left corner in the first sum is zero. Using (8.19), it is easily verified that the lower left corner in the third sum is also zero.

The lower left corner in the second sum in (8.21) is equal to

$$\sum_{j=u+1}^v t_{ij} \eta_j \bar{t}_{kj},$$

where t_{ij} is the entry on the main diagonal of T_{ij} . Thus,

$$\delta_{ik} \tilde{\eta}_i = \sum_{j=u+1}^v t_{ij} \eta_j \bar{t}_{kj}.$$

It follows that

$$\tilde{\eta}_{u+1} \oplus \dots \oplus \tilde{\eta}_v = [t_{ik}]_{i,k=u+1}^v (\eta_{u+1} \oplus \dots \oplus \eta_v) \left([t_{ik}]_{i,k=u+1}^v \right)^*. \quad (8.22)$$

Since the left hand side of (8.22) is invertible, the matrix $[t_{ik}]_{i,k=u+1}^v$ is also invertible. Now Theorem 4.1 guarantees that the two systems $\{\tilde{\eta}_{u+1}, \dots, \tilde{\eta}_v\}$ and $\{\eta_{u+1}, \dots, \eta_v\}$ have the same number of +1's (and also the same number of -1's). Thus, within each set of blocks of equal size ℓ_j , the number of η_j 's which are equal to +1 (resp. to -1) coincides with the number of $\tilde{\eta}_j$'s which are equal to +1 (resp. to -1). This shows that (8.17) is indeed obtained from (8.16) after a permutation of blocks.

Finally, assume that $A + \lambda B$ is congruent to (8.15) with possibly different signs $\tilde{\delta}_j$. Arguing as in the preceding case, we obtain the equalities

$$T_{ij} (\delta_j F_{k_j}) = (\tilde{\delta}_i F_{k_i}) S_{ij}, \quad T_{ij} (\delta_j G_{k_j}) = (\delta_i G_{k_i}) S_{ij}.$$

The proof that the form

$$\bigoplus_{j=1}^r \tilde{\delta}_j (F_{k_j} + G_{k_j})$$

is obtained from (8.15) after a permutation of blocks, proceeds from now on in the same way (letting $\alpha = 0$) as the proof that (8.16) and (8.17) are the same up to a permutation of blocks.

The proof of the uniqueness part of Theorem 6.1 is complete.

9 Pairs of real symmetric matrices

The results of Section 5–8 have analogues for *real symmetric* matrix pencils $A + \lambda B$. Here, A and B are real symmetric $n \times n$ matrices. Mostly, the proofs are very similar to the corresponding proofs of Sections 5–8. The results on canonical forms are stated in this section, and the necessary changes in the proofs are indicated. Let us begin with the canonical form under strict equivalence.

It will be helpful to introduce another standard $2m \times 2m$ matrix

$$H_{2m} = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 \\ 0 & & & 0 & -1 \\ \vdots & & 1 & 0 & \\ & & 0 & -1 & \\ & & & & \vdots \\ 1 & 0 & & & 0 \\ 0 & -1 & \cdots & 0 & 0 \end{bmatrix}.$$

Clearly, H_{2m} is real symmetric. It is easy to verify the following connection with real Jordan blocks and sip matrices:

$$(\lambda I_{2m} + J_{2m}(\mu \pm i\nu)) F_{2m} = (\lambda + \mu)F_{2m} + \nu H_{2m} + \begin{bmatrix} F_{2m-2} & 0 \\ 0 & 0_2 \end{bmatrix}. \quad (9.1)$$

Using this equality in place of (5.5), and arguing as in Section 5, we obtain the canonical form under strict equivalence:

Theorem 9.1 *Every real symmetric matrix pencil $A + \lambda B$ is strictly equivalent to a real symmetric matrix pencil of the form*

$$\begin{aligned} 0_{u \times u} \oplus & \bigoplus_{j=1}^p \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_j} \\ 0 & 0 & 0 \\ F_{\varepsilon_j} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_j+1} \right) \\ & \oplus \bigoplus_{j=1}^r (F_{k_j} + \lambda G_{k_j}) \oplus \bigoplus_{j=1}^q ((\lambda + \alpha_j)F_{\ell_j} + G_{\ell_j}) \oplus \\ & \oplus \bigoplus_{j=1}^s \left((\lambda + \mu_j)F_{2m_j} + \nu_j H_{2m_j} + \begin{bmatrix} F_{2m_j-2} & 0 \\ 0 & 0_2 \end{bmatrix} \right), \end{aligned} \quad (9.2)$$

where $\varepsilon_1 \leq \cdots \leq \varepsilon_p$ and $k_1 \leq \cdots \leq k_r$ are positive integers, α_j are real numbers, and $\mu_j + i\nu_j$ are nonreal numbers ($\mu_j, \nu_j \in \mathbf{R}$).

The form (9.2) is uniquely determined by $A + \lambda B$ up to a permutation of the blocks $(\lambda + \alpha_j)F_{\ell_j} + G_{\ell_j}$ ($j = 1, \dots, q$), and up to a permutation of the blocks

$$(\lambda + \mu_j)F_{2m_j} + \nu_j H_{2m_j} + \begin{bmatrix} F_{2m_j-2} & 0 \\ 0 & 0_2 \end{bmatrix} \quad (j = 1, \dots, s), \quad (9.3)$$

with possible replacement of ν_j by $-\nu_j$ in each of the blocks (9.3).

Again, using \widetilde{G}_m in place of G_m and $\begin{bmatrix} 0_2 & 0 \\ 0 & F_{2m-2} \end{bmatrix}$ in place of $\begin{bmatrix} F_{2m-2} & 0 \\ 0 & 0_2 \end{bmatrix}$, an alternative form of the canonical form can be obtained.

We say that two real symmetric $n \times n$ matrix pencils $A_1 + \lambda B_1$ and $A_2 + \lambda B_2$ are *congruent over the real numbers*, in short, *R-congruent*, if

$$P(A_1 + \lambda B_1)P^* = A_2 + \lambda B_2$$

for some invertible real matrix P . The real analogue of Theorem 6.1 runs as follows:

Theorem 9.2 *Every real symmetric matrix pencil $A + \lambda B$ is R-congruent to a real symmetric matrix pencil of the form*

$$\begin{aligned} 0_{u \times u} \oplus & \bigoplus_{j=1}^p \left(\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_j} \\ 0 & 0_1 & 0 \\ F_{\varepsilon_j} & 0 & 0 \end{bmatrix} + G_{2\varepsilon_j+1} \right) \\ & \oplus \bigoplus_{j=1}^r (\delta_j (F_{k_j} + \lambda G_{k_j})) \oplus \bigoplus_{j=1}^q (\eta_j ((\lambda + \alpha_j) F_{\ell_j} + G_{\ell_j})) \\ & \oplus \bigoplus_{j=1}^s \left((\lambda + \mu_j) F_{2m_j} + \nu_j H_{2m_j} + \begin{bmatrix} F_{2m_j-2} & 0 \\ 0 & 0_2 \end{bmatrix} \right), \end{aligned} \quad (9.4)$$

where $\varepsilon_j, k_j, \alpha_j, \mu_j + i\nu_j$ are as in Theorem 9.1, and $\delta_1, \dots, \delta_r, \eta_1, \dots, \eta_q$ are signs $+1$ or -1 .

The form (9.4) is uniquely determined by $A + \lambda B$ up to a permutation of the blocks $\delta_j (F_{k_j} + \lambda G_{k_j})$ ($j = 1, \dots, r$), a permutation of the blocks $\eta_j ((\lambda + \alpha_j) F_{\ell_j} + G_{\ell_j})$, ($j = 1, \dots, q$) and a permutation of the blocks

$$(\lambda + \mu_j) F_{2m_j} + \nu_j H_{2m_j} + \begin{bmatrix} F_{2m_j-2} & 0 \\ 0 & 0_2 \end{bmatrix} \quad j = 1, \dots, s,$$

with possible replacement of ν_j by $-\nu_j$ within each such block.

As with Theorem 6.1, the result of Theorem 9.2 has a long history. For the nonsingular case, it was proved by Muth [41], and later by Trott [53]. In the general case it was proved by Williamson [60]. Again, the result has been rediscovered and extended in many ways, and the reader is referred to the extensive bibliography of [52].

Our proof of Theorem 9.2 follows the same pattern as that of Theorem 6.1. We indicate below the changes that must be made to that argument.

The proof of existence of the form (9.4) is essentially the same as the proof used in Section 7. The only differences are that now Theorem 9.1 is used instead of Theorem

5.1 to obtain the formula (7.2) with real invertible matrices P and Q ; instead of Z_2 of (7.7) consider

$$Z_{2m}(\lambda, \mu, \nu) := (\lambda + \mu)F_{2m} + \nu H_{2m} + \begin{bmatrix} F_{2m-2} & 0 \\ 0 & 0_2 \end{bmatrix} \quad (9.5)$$

of equation (9.1). (Here μ, ν are real and $\nu \neq 0$). It is found that

$$\begin{bmatrix} K & 0 & \dots & 0 \\ 0 & K & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K \end{bmatrix} Z_{2m}(\lambda, \mu, \nu) \begin{bmatrix} K^T & 0 & \dots & 0 \\ 0 & K^T & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & K^T \end{bmatrix} = -Z_{2m}(\lambda, \mu, \nu),$$

where $K = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Finally, the real Jordan form (Theorem 2.2) is used in place of (7.8).

As for the proof of uniqueness of (9.4), the arguments of Section 8 are followed. Note that Lemma 8.1 and its proof also hold for real symmetric matrix pencils (the only difference is that now we use the existence part of Theorem 9.2 instead of the existence part of Theorem 6.1). Once Lemma 8.1 is proved for real symmetric matrix pencils, the proof of uniqueness of (9.4) is the same as in the proof in Section 8 (Theorem 9.1 is used in place of Theorem 5.1).

Note that the results of Corollaries 6.2 and 6.3 and their proofs extend verbatim to the real case.

10 Simultaneous diagonalization

It is a well-known fact that if A and B are two hermitian matrices, and if one of them is positive definite (or more generally, there is a positive definite linear combination of A and B with real coefficients), then A and B are simultaneously diagonalizable by congruence, i.e., there exist real diagonal matrices D_1 and D_2 such that the hermitian pencil $A + \lambda B$ is congruent (\mathbf{R} -congruent if both A and B are real) to $D_1 + \lambda D_2$. A proof of this fact is given in several texts including Gantmacher and Krein [13], Gantmacher [12], and Franklin [11]. The result is easily obtained by inspection of Theorems 6.1 and 9.2, and it generally fails if the positive definiteness hypothesis is omitted.

Theorems 6.1 and 9.2 lead to more refined results concerning simultaneous diagonalizability with the positive definiteness hypothesis relaxed.

Theorem 10.1 *Let $A, B \in \mathbf{C}^{n \times n}$ be hermitian matrices. Assume that there exists a linear combination*

$$C = \alpha A + \beta B, \quad \alpha, \beta \in \mathbf{R},$$

such that C is positive semidefinite, and

$$\text{Ker}(C) \subseteq \text{Ker}(A) \cap \text{Ker}(B). \quad (10.1)$$

Then A and B are simultaneously diagonalizable by congruence or, if both A and B are real, by \mathbb{R} -congruence.

In particular, the hypotheses of Theorem 10.1 are satisfied if C is positive definite.

Proof. Consider the complex case (the proof in the real case is virtually identical). We may clearly assume that at least one of α and β is nonzero (otherwise the theorem is trivial). By interchanging the roles of A and B and scaling (if necessary) we may assume that $\beta = 1$ and, furthermore, by replacing B with $B + \alpha A$, we may further assume that $C = B$. Finally, take $A + \lambda B$ in the form (6.1). The condition that B is positive semidefinite easily implies that

$$A + \lambda B = 0_{u \times u} \oplus \delta_1 (F_{k_1} + \lambda G_{k_1}) \oplus \cdots \oplus \delta_r (F_{k_r} + \lambda G_{k_r}) \oplus \text{diag}(\lambda + \alpha_1, \dots, \lambda + \alpha_q), \quad (10.2)$$

in the notation of Theorem 6.1, where $k_j \leq 2$. If there is a term in (10.2) with $k_j = 2$, then $\text{Ker}(B)$ is not contained in $\text{Ker}(A)$, a contradiction with (10.1). Thus, we must have all $k_j = 1$, and the righthand side of (10.2) is diagonal. \square

The conditions of Theorem 10.1 are obviously not necessary for simultaneous diagonalization of a pair of hermitian matrices. A criterion for simultaneous diagonalization is given by Rao and Mitra [47, Chapter 6]. Connections between simultaneous diagonalization and positive semidefinite linear combinations (a topic of the next section) were also explored by Hestenes [18] and Au-Yeung [3].

11 Positive semidefinite linear combinations

If A and B are hermitian and there are real numbers α and β , not both zero, such that $\alpha A + \beta B$ is positive semidefinite, we say that (A, B) admits a *positive semidefinite linear combination*. Similarly, if there is a positive definite real linear combination of A and B we say that (A, B) admits a *positive definite linear combination*. Theorem 10.1 concerns a role for this property in the study of simultaneous diagonalization. However, these concepts arise more frequently and, in this section, Theorems 6.1 and 9.1 are used in the discussion of several characterizations of these two properties (Theorems 11.1 and 11.3 in the semidefinite case which seem to be new, and Theorem 11.5 in the definite case).

Theorem 11.1 *Let $A, B \in \mathbb{C}^{n \times n}$ be hermitian. Then the pair (A, B) admits a positive semidefinite linear combination if and only if the following property holds:*

(α) *For any $x \in \mathbb{C}^n$ such that $(Ax, x) = (Bx, x) = 0$, the two vectors Ax and Bx are \mathbb{R} -linearly dependent.*

Other characterizations of the positive semidefinite linear combination property, and an alternative proof of Theorem 11.1, are given by Cheung et al. ([7]), in the context of selfadjoint operators on Hilbert spaces. See also Lancaster and Ye, [33].

For the proof of Theorem 11.1, it will be convenient to dispose of a particular case first.

Lemma 11.2 *Let $A, B \in \mathbb{C}^{n \times n}$ be hermitian with block forms*

$$A = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}, \quad B = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix},$$

where $D_1 \in \mathbb{C}^{p \times p}$ and $D_2 \in \mathbb{C}^{q \times q}$ are hermitian. If the property (α) is satisfied, then the pair (A, B) admits a positive semidefinite linear combination.

Proof. Applying unitary similarity to each of D_1 and D_2 , we may assume that D_1 and D_2 are diagonal. The proof proceeds by induction on p and q . The basis of induction, i.e., $p = 0$ or $q = 0$ is trivial (then $\pm B$ is positive definite). The case $p = q = 1$ can be handled by elementary considerations, and we omit the details here. Suppose at least one of p and q is larger than 1; say $p > 1$. We may assume that the largest eigenvalue of D_1 , call it γ , is in the top left corner of D_1 . Replace A with $A - \gamma B$. This transformation does not change the property (α) and the property of having a (nontrivial) positive semidefinite linear combination. Then

$$A = \begin{bmatrix} 0_{1 \times 1} & 0 & 0 \\ 0 & \tilde{D}_1 & 0 \\ 0 & 0 & \tilde{D}_2 \end{bmatrix},$$

where \tilde{D}_1 is negative semidefinite. Applying the induction hypothesis to

$$\tilde{A} := \begin{bmatrix} \tilde{D}_1 & 0 \\ 0 & \tilde{D}_2 \end{bmatrix}, \quad \tilde{B} := \begin{bmatrix} I_{p-1} & 0 \\ 0 & -I_q \end{bmatrix},$$

we see that there exist $\alpha, \beta \in \mathbb{R}$ not both zeros such that $\alpha \tilde{A} + \beta \tilde{B}$ is positive semidefinite. Since \tilde{D}_1 is negative semidefinite, it follows that $\beta \geq 0$. But then $\alpha A + \beta B$ is positive semidefinite as well. □

Proof of Theorem 11.1. Assume $\alpha A + \beta B$ is positive semidefinite for some $\alpha, \beta \in \mathbb{R}$, not both zero. If $(Ax, x) = (Bx, x) = 0$, then $((\alpha A + \beta B)x, x) = 0$, and in view of positive semidefiniteness of $\alpha A + \beta B$ we have $(\alpha A + \beta B)x = 0$. So the \mathbb{R} -linear dependence of Ax and Bx follows.

Conversely, assume that (α) holds. Without loss of generality it may be assumed that the hermitian pencil $A + \lambda B$ has the form (6.1). Clearly, condition (α) holds true for every constituent diagonal block in the direct sum (6.1).

Now consider whether condition (α) holds for each of the possible diagonal blocks in (6.1). Clearly, (α) holds for $0_{u \times u}$.

Let

$$A' + \lambda B' = \lambda \begin{bmatrix} 0 & 0 & F_\varepsilon \\ 0 & 0 & 0 \\ F_\varepsilon & 0 & 0 \end{bmatrix} + G_{2\varepsilon+1}$$

where F_ε is a sip matrix, and let $x = e_1 \in \mathbb{C}^n$, the first unit coordinate vector. Then $(A'x, x) = (B'x, x) = 0$, but $A'x$ and $B'x$ are linearly independent. Thus, (α) does not hold for the pair (A', B') .

Let

$$A' + \lambda B' = F_k + \lambda G_k, \quad \text{or} \quad A' + \lambda B' = (\lambda + \alpha)F_k + G_k, \quad (\alpha \in \mathbb{R}).$$

Then, using the same $x = e_1$, it will be shown that (α) does not hold for the pair (A', B') , unless $k \leq 2$.

Let

$$A' + \lambda B' = \begin{bmatrix} 0 & (\lambda + \beta)F_m + G_m \\ (\lambda + \bar{\beta})F_m + G_m & 0 \end{bmatrix}, \quad \beta \in \mathbb{C} \setminus \mathbb{R}.$$

Then, using $x = e_1$ we see that (α) does not hold for (A', B') .

Next, consider the following four possibilities (11.1) - (11.4) for the pair (A', B') . It will be proved that, in each of the four cases, the pair (A', B') does not have property (α) :

$$A' = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \oplus \pm \begin{bmatrix} 1 & \alpha \\ \alpha & 0 \end{bmatrix}, \quad B' = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \oplus \pm \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \alpha \in \mathbb{R}. \quad (11.1)$$

$$A' = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \quad B' = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \alpha \in \mathbb{R}. \quad (11.2)$$

$$A' = \begin{bmatrix} 1 & \alpha \\ \alpha & 0 \end{bmatrix} \oplus \begin{bmatrix} -1 & -\alpha \\ -\alpha & 0 \end{bmatrix}, \quad B' = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \quad \alpha \in \mathbb{R}. \quad (11.3)$$

$$A' = \begin{bmatrix} 1 & \alpha \\ \alpha & 0 \end{bmatrix} \oplus \pm \begin{bmatrix} 1 & \beta \\ \beta & 0 \end{bmatrix}, \quad B' = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \oplus \pm \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \alpha, \beta \in \mathbb{R}, \quad \alpha \neq \beta. \quad (11.4)$$

Now, in each of the four cases, a vector $x \in \mathbb{R}^4$ is formulated for which $(Ax, x) = (Bx, x) = 0$, and Ax, Bx are \mathbb{R} -linearly independent:

$$x = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{if } (A', B') \text{ are given by (11.1) or (11.4),}$$

and

$$x = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{if } (A', B') \text{ are given by (11.2) or (11.3).}$$

Using the above analysis and, if necessary, replacing (A, B) by $(-A, -B)$, we see that the pair (A, B) can be taken in one of the forms

$$\begin{aligned} A + \lambda B &= 0_{u \times u} \oplus \delta_1 \oplus \cdots \oplus \delta_r \oplus \eta_1 (\lambda + \alpha_1) \oplus \cdots \oplus \eta_q (\lambda + \alpha_q) \\ &\oplus \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) \oplus \cdots \oplus \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right), \end{aligned} \quad (11.5)$$

where δ_j and η_j are signs ± 1 , and $\alpha_j \in \mathbf{R}$, or

$$\begin{aligned} A + \lambda B &= 0_{u \times u} \oplus \delta_1 \oplus \cdots \oplus \delta_r \oplus \eta_1 (\lambda + \alpha_1) \oplus \cdots \oplus \eta_q (\lambda + \alpha_q) \\ &\oplus \left(\begin{bmatrix} 1 & \gamma \\ \gamma & 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) \oplus \cdots \oplus \left(\begin{bmatrix} 1 & \gamma \\ \gamma & 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right), \end{aligned} \quad (11.6)$$

where δ_j and η_j are signs ± 1 , and $\alpha_j, \gamma \in \mathbf{R}$. Without loss of generality it may be assumed that $u = 0$, i.e., the term $0_{u \times u}$ does not appear in (11.5) and (11.6).

Suppose first that (11.5) holds, and assume that at least one of the blocks $\eta_j (\lambda + \alpha_j)$ and at least one of the blocks $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ are present. If one of the signs η_j is -1 , say $\eta_1 = -1$, then we obtain a contradiction with the hypothesis that (A, B) satisfies the property (α) , because the pair

$$(A', B') = \left(\begin{bmatrix} q & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right), \quad q \in \mathbf{R}, \quad (11.7)$$

does not satisfy the property (α) , as can be easily seen by considering the vector $x = [1 \ 1 \ -q/2]^T$. Indeed, $(A'x, x) = (B'x, x) = 0$, but $A'x = [q \ -q/2 \ 1]^T$ and $B'x = [-1 \ 1 \ 0]^T$ are linearly independent. Thus, all $\eta_j = 1$. But then B is positive semidefinite, and we are done in this case.

If no blocks $\eta_j (\lambda + \alpha_j)$ are present, then B is positive semidefinite, and it may be assumed that no blocks $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ are present. We may further assume (adding if necessary a suitable scalar multiple of B to A), that all the α_j 's are different from zero. Now the result follows from Lemma 11.2 (with the roles of A and B interchanged), upon applying a suitable diagonal congruence to A and B .

Next, suppose that (11.6) holds. Then

$$\begin{aligned}
B + \lambda(A - \gamma B) &= \lambda\delta_1 \oplus \cdots \oplus \lambda\delta_r \\
&\oplus (\eta_1 + \lambda(\eta_1\alpha_1 - \eta_1\gamma)) \oplus \cdots \oplus (\eta_q + \lambda(\eta_q\alpha_q - \eta_q\gamma)) \\
&\oplus \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) \oplus \cdots \oplus \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right).
\end{aligned} \tag{11.8}$$

This form can be easily reduced, after a suitable diagonal congruence, to (11.5), to complete the proof. \square

The real analogue of Theorem 11.1 reads as follows:

Theorem 11.3 *Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric. Assume that the pencil $A + \lambda B$ is not \mathbb{R} -congruent to a pencil of the form*

$$0_{u \times u} \oplus \begin{bmatrix} \nu & \mu \\ \mu & -\nu \end{bmatrix} + \lambda \left(0_{u \times u} \oplus \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right), \quad \mu, \nu \in \mathbb{R}, \quad \nu \neq 0, \tag{11.9}$$

where u is a nonnegative integer. Then the pair (A, B) admits a positive semidefinite linear combination if and only if the following property holds:

(α') *For any $x \in \mathbb{R}^n$ for which $(Ax, x) = (Bx, x) = 0$, the two vectors Ax and Bx are \mathbb{R} -linearly dependent.*

In case $A + \lambda B$ is \mathbb{R} -congruent to a pencil of the form (11.9), one easily verifies that (α') is satisfied, but the pair (A, B) does not admit a (nontrivial) positive semidefinite linear combination.

Proof. Suppose the pair (A, B) satisfies the property (α'). By Theorem 9.1 we may assume that $A + \lambda B$ is given by the right-hand side of (9.2). As in the proof of

Theorem 11.1, we assume that $u = 0$, and verify that no blocks $\lambda \begin{bmatrix} 0 & 0 & F_{\varepsilon_j} \\ 0 & 0 & 0 \\ F_{\varepsilon_j} & 0 & 0 \end{bmatrix} +$

$G_{2\varepsilon_j+1}$ are present, no blocks $\delta_j(F_{k_j} + \lambda G_{k_j})$ with $k_j \geq 3$ are present, and no blocks $\eta_j((\lambda + \alpha_j)F_{\ell_j} + G_{\ell_j})$ with $\ell_j \geq 3$ are present. Furthermore, the pairs of blocks (11.1), (11.2), (11.3), and (11.4), as well as their negatives, do not have property (α'), and therefore cannot be present in (9.2).

Consider the block

$$A' + \lambda B' = (\lambda + \mu)F_{2m} + \nu H_{2m} + \begin{bmatrix} F_{2m-2} & 0 \\ 0 & 0_2 \end{bmatrix}, \tag{11.10}$$

where $\mu, \nu \in \mathbb{R}$ and $\nu \neq 0$. If $m > 1$, then for $x = e_1 \in \mathbb{R}^{2m}$ we have $(A'x, x) = (B'x, x) = 0$ and $A'x, B'x$ are linearly independent, a contradiction with property (α') . If we have a pair of blocks of the type (11.10) of size two,

$$B' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad A' = \begin{bmatrix} \nu_1 & \mu_1 & 0 & 0 \\ \mu_1 & -\nu_1 & 0 & 0 \\ 0 & 0 & \nu_2 & \mu_2 \\ 0 & 0 & \mu_2 & -\nu_2 \end{bmatrix}, \quad (11.11)$$

then, assuming without loss of generality that $\nu_1 < 0 < \nu_2$ (see Theorem 9.2), and letting $p, q \in \mathbb{R} \setminus \{0\}$ be such that $\nu_1 p^2 + \nu_2 q^2 = 0$, it can be verified that the vector $x = [p \ 0 \ q \ 0]^T$ satisfies $(A'x, x) = (B'x, x) = 0$ and $A'x$ and $B'x$ are linearly independent. Thus, a pair of blocks (11.11) cannot be present in (9.2).

Summarizing the above analysis, we see that the pair (A, B) can be taken in one of the forms

$$\begin{aligned} A + \lambda B &= \delta_1 \oplus \cdots \oplus \delta_r \oplus \eta_1 (\lambda + \alpha_1) \oplus \cdots \oplus \eta_q (\lambda + \alpha_q) \\ &\oplus \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) \oplus \cdots \oplus \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) \\ &\oplus ((\lambda + \mu)F_2 + \nu H_2), \end{aligned} \quad (11.12)$$

where δ_j and η_j are signs ± 1 , $\alpha_j, \mu, \nu \in \mathbb{R}$, and $\nu \neq 0$, or

$$\begin{aligned} A + \lambda B &= \delta_1 \oplus \cdots \oplus \delta_r \oplus \eta_1 (\lambda + \alpha_1) \oplus \cdots \oplus \eta_q (\lambda + \alpha_q) \\ &\oplus \left(\begin{bmatrix} 1 & \gamma \\ \gamma & 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) \oplus \cdots \oplus \left(\begin{bmatrix} 1 & \gamma \\ \gamma & 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) \\ &\oplus ((\lambda + \mu)F_2 + \nu H_2), \end{aligned} \quad (11.13)$$

where δ_j and η_j are signs ± 1 , $\alpha_j, \gamma, \mu, \nu \in \mathbb{R}$, and $\nu \neq 0$. It may be assumed that the block $(\lambda + \mu)F_2 + \nu H_2$ is present in (11.12) and (11.13), otherwise the argument can be completed using the proof of Theorem 11.1.

To complete the proof it only remains to verify that the following four pairs of matrices (A', B') given by (11.14) - (11.17) below, do not have the property (α') :

$$A' = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \nu & \mu \\ 0 & \mu & -\nu \end{bmatrix}, \quad B' = \begin{bmatrix} \pm 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \alpha, \mu, \nu \in \mathbb{R}, \quad \nu \neq 0, \quad (11.14)$$

$$A' = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \nu & \mu \\ 0 & \mu & -\nu \end{bmatrix}, \quad B' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \alpha, \mu, \nu \in \mathbb{R}, \quad \nu \neq 0, \quad \alpha \neq 0, \quad (11.15)$$

$$A' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \nu & \mu \\ 0 & 0 & \mu & -\nu \end{bmatrix}, \quad B' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mu, \nu \in \mathbb{R}, \quad \nu \neq 0, \quad (11.16)$$

$$A' = \begin{bmatrix} 1 & \gamma & 0 & 0 \\ \gamma & 0 & 0 & 0 \\ 0 & 0 & \nu & \mu \\ 0 & 0 & \mu & -\nu \end{bmatrix}, \quad B' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \gamma, \mu, \nu \in \mathbb{R}, \quad \nu \neq 0. \quad (11.17)$$

Consider (11.14) and (11.15) together, and let $\kappa = \pm 1$ or $\kappa = 0$ be the top left corner of B' . Define the real numbers p and q by

$$(p + qi)^2 = -\frac{\alpha - \kappa\mu}{\nu} - \kappa i, \quad i = \sqrt{-1}.$$

Since the case $\kappa = \alpha = 0$ is excluded, at least one of the numbers p and q is nonzero. Furthermore,

$$2pq + \kappa = 0, \quad (p^2 - q^2)\nu + \alpha - \kappa\mu = 0,$$

and therefore the vector $x = [1 \ p \ q]^T$ satisfies $(A'x, x) = (B'x, x) = 0$. On the other hand,

$$A'x = \begin{bmatrix} \alpha \\ p\nu + q\mu \\ -q\nu + p\mu \end{bmatrix}, \quad B'x = \begin{bmatrix} \kappa \\ q \\ p \end{bmatrix}$$

are linearly independent. For the pair (11.16), define $p, q \in \mathbb{R}$ by $(p + qi)^2 = \frac{\mu}{\nu} - i$, and verify using the vector $x = [1 \ 0 \ p \ q]^T$ that (α') does not hold for (11.16). For the pair (11.17) we argue similarly using the vector $x = [1 \ 1 \ p \ q]^T$, where $p, q \in \mathbb{R}$ are defined by $(p + qi)^2 = (2\mu - 1 - 2\gamma)\nu^{-1} - 2i$. □

Corollary 11.4 *Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric matrices such that $\det(A + \lambda B)$ is not identically zero, and $n \neq 2$. Then the pair (A, B) admits a positive semidefinite linear combination if and only if the property (α') holds.*

The following result provides a criterion for the existence of a positive definite linear combination. It was proved by Finsler [10] and later reproved many times using various methods; by Au-Yeung [2], for example, and see the survey by Uhlig [56]. Note also that the critical condition (11.18) appears in the definition of a class of operators in spaces with indefinite inner product (generally of infinite dimension) known as *Pesonen operators*, [42].

Theorem 11.5 (a) *Let $A, B \in \mathbb{C}^{n \times n}$ be hermitian. Then the pair (A, B) admits a positive definite linear combination if and only if the following property holds:*

$$(Ax, x) = (Bx, x) = 0, \quad x \in \mathbb{C}^n \quad \implies \quad x = 0. \quad (11.18)$$

(b) *Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric. Assume $n \neq 2$. Then the pair (A, B) admits a positive definite linear combination if and only if the following property holds:*

$$(Ax, x) = (Bx, x) = 0, \quad x \in \mathbb{R}^n \quad \implies \quad x = 0.$$

A proof of Theorem 11.5 can be obtained using the canonical forms of Theorems 6.1 and 9.2. However, there is a much easier proof by Au-Yeung and Poon (in [4]). This is based on the convexity property of the numerical range of the matrix $A + iB$ in the complex case, and on the convexity property of the real joint numerical range

$$W(A, B) := \{((Ax, x), (Bx, x)) \in \mathbb{R}^2 : x \in \mathbb{R}^n, \|x\| = 1\}$$

in the real case (if $n \geq 3$). Convexity of the numerical ranges of complex matrices is the celebrated Toeplitz-Hausdorff theorem. Convexity of real joint numerical ranges for pairs of real symmetric matrices of sizes larger than 2 is also well-known. (This is a result of Brinkman [6]; see also a paper of Marcus, [38]).

12 Canonical forms of H -selfadjoint matrices

Theorems 6.1 and 9.2 immediately lead to canonical forms for matrices which are selfadjoint with respect to an indefinite inner product. This applies in both the real and the complex cases. Attention is confined to *non-degenerate* inner products (when H , below, is nonsingular).

First consider the complex case. If H is an invertible hermitian $n \times n$ matrix, then an indefinite inner product $[\cdot, \cdot]$ is defined in \mathbb{C}^n by

$$[x, y] = (Hx, y).$$

Recall that a matrix $A \in \mathbb{C}^{n \times n}$ is called *selfadjoint* with respect to the indefinite scalar product $[\cdot, \cdot]$, or, in short, is *H -selfadjoint*, if $[Ax, y] = [x, Ay]$ for every $x, y \in \mathbb{C}^n$. It is easily seen that A is H -selfadjoint if and only if $HA = A^*H = (HA)^*$. Thus, Theorem 6.1 can be applied to the hermitian pencil $HA + \lambda H$. Denoting the canonical form of $HA + \lambda H$ under congruence by $C_1 + \lambda C_2$, we have

$$HA = S^*C_1S, \quad H = S^*C_2S$$

for some invertible matrix S . Solving for A , we obtain:

$$A = H^{-1}S^*C_1S = S^{-1}C_2^{-1}C_1S.$$

Thus, the pair $(C_2^{-1}C_1, C_2)$ can serve as the canonical form of (A, H) under the *congruence similarity*: $(A, H) \rightarrow (S^{-1}AS, S^*HS)$ for some invertible matrix $S \in \mathbb{C}^{n \times n}$. Note that here H (and therefore S), is invertible. Note also that, in the notation of Theorem 6.1,

$$F_{\ell_j}^{-1}(\alpha_j F_{\ell_j} + G_{\ell_j}) = J_{\ell_j}(\alpha_j)$$

and

$$\begin{bmatrix} 0 & F_{m_j} \\ F_{m_j} & 0 \end{bmatrix}^{-1} \left(\begin{bmatrix} 0 & \beta_j F_{m_j} \\ \bar{\beta}_j F_{m_j} & 0 \end{bmatrix} + \begin{bmatrix} 0 & G_{m_j} \\ G_{m_j} & 0 \end{bmatrix} \right) = \begin{bmatrix} J_{m_j}(\beta_j) & 0 \\ 0 & J_{m_j}(\bar{\beta}_j) \end{bmatrix},$$

i.e. the canonical matrix $C_2^{-1}C_1$ is simply the Jordan form of A . The following result is obtained:

Theorem 12.1 *Let $A \in \mathbb{C}^{n \times n}$ be H -selfadjoint. Then there exists an invertible matrix $S \in \mathbb{C}^{n \times n}$ such that*

$$S^{-1}AS = J_{\ell_1}(\alpha_1) \oplus \cdots \oplus J_{\ell_q}(\alpha_q) \oplus \left[\begin{array}{cc} J_{m_1}(\beta_1) & 0 \\ 0 & J_{m_1}(\bar{\beta}_1) \end{array} \right] \oplus \cdots \oplus \left[\begin{array}{cc} J_{m_s}(\beta_s) & 0 \\ 0 & J_{m_s}(\bar{\beta}_s) \end{array} \right] \quad (12.1)$$

$$S^*HS = \eta_1 F_{\ell_1} \oplus \cdots \oplus \eta_q F_{\ell_q} \oplus F_{2m_1} \oplus \cdots \oplus F_{2m_s}, \quad (12.2)$$

where $\alpha_1, \dots, \alpha_q$ are (not necessarily distinct) real numbers, β_1, \dots, β_s are (not necessarily distinct) complex numbers with positive imaginary parts, and $\eta_1, \dots, \eta_q \in \{1, -1\}$.

The form (12.1), (12.2) is uniquely determined by A and H , up to a permutation of the blocks $\{J_{\ell_j}(\alpha_j)\}_{j=1}^q$ and the same simultaneous permutation of the blocks

$\{\eta_j F_{\ell_j}\}_{j=1}^q$, and up to a permutation of the blocks $\left\{ \left[\begin{array}{cc} J_{m_j}(\beta_j) & 0 \\ 0 & J_{m_j}(\bar{\beta}_j) \end{array} \right] \right\}_{j=1}^s$ and the same simultaneous permutation of the blocks $\{F_{2m_j}\}_{j=1}^s$.

For the real case, Theorem 9.2 determines the canonical form of an H -selfadjoint real matrix A (where H is an invertible real symmetric matrix) under real congruence similarity.

Theorem 12.2 *Let $H \in \mathbb{R}^{n \times n}$ be an invertible symmetric matrix, and let $A \in \mathbb{R}^{n \times n}$ be H -selfadjoint. Then there exists an invertible real matrix S such that*

$$S^{-1}AS = J_{\ell_1}(\alpha_1) \oplus \cdots \oplus J_{\ell_q}(\alpha_q) \oplus J_{2m_1}(\mu_1 \pm i\nu_1) \oplus \cdots \oplus J_{2m_s}(\mu_s \pm i\nu_s), \quad (12.3)$$

$$S^*HS = \eta_1 F_{\ell_1} \oplus \cdots \oplus \eta_q F_{\ell_q} \oplus F_{2m_1} \oplus \cdots \oplus F_{2m_s}, \quad (12.4)$$

where $\alpha_1, \dots, \alpha_q \in \mathbb{R}$; $\mu_1 \pm i\nu_1, \dots, \mu_s \pm i\nu_s$ are pairs of complex conjugate numbers with positive ν_1, \dots, ν_s , and $\eta_1, \dots, \eta_q \in \{1, -1\}$.

The form (12.3), (12.4) is uniquely determined by A and H , up to a permutation of pairs of blocks $\{(J_{\ell_j}(\alpha_j), \eta_j F_{\ell_j})\}_{j=1}^q$ and up to a permutation of pairs of blocks $\{(J_{2m_j}(\mu_j \pm i\nu_j), F_{2m_j})\}_{j=1}^s$.

Theorems 12.1 and 12.2 are key results in the theory and applications of linear transformations that are selfadjoint with respect to an indefinite inner product. (See the monograph of Krein [25] and the more recent books of Azizov et al. [1], Iohvidov et al. [23], and Gohberg et al. [16], and, for the approach via canonical pairs, see the works of Weierstrass [62], Kronecker [28], and Malcev [37]).

In view of its importance, particularly in the factorization theory of matrix polynomials with hermitian coefficients (initiated in the works of Gohberg et al [14], and also [15], [16]) and of symmetric rational matrix functions studied by Ran [44], the present authors [29], and Ran and Rodman [45], [46]), a further basic result based on Theorem 12.1 is presented:

Theorem 12.3 *Let $A \in \mathbb{C}^{n \times n}$ be H -selfadjoint. Then there exist:*

- (i) *An A -invariant subspace $\mathcal{M}_+ \subseteq \mathbb{C}^n$ of dimension $\text{In}_+(H)$ such that $[x, x] \geq 0$ for every $x \in \mathcal{M}_+$;*
- (ii) *An A -invariant subspace $\mathcal{M}_- \subseteq \mathbb{C}^n$ of dimension $\text{In}_-(H)$ such that $[x, x] \leq 0$ for every $x \in \mathcal{M}_-$.*

Moreover, the subspaces \mathcal{M}_\pm can be chosen so that all eigenvalues of the restriction $A_{\mathcal{M}_\pm}$ have nonnegative imaginary parts, or, alternatively, all eigenvalues of the restriction $A_{\mathcal{M}_\pm}$ have nonpositive imaginary parts.

This result is, in fact, a particular case of more general theorems originating with Pontryagin [43]. They assert the existence of invariant maximal semidefinite subspaces for certain classes of operators on infinite dimensional spaces with an indefinite inner product. For more recent developments of this kind see the monographs of Azizov et al. [1], and Iohvidov et al. [23].

Proof. The existence of \mathcal{M}_+ as required in (i) will be proved, with the further property that all eigenvalues of $A_{\mathcal{M}_+}$ have nonnegative imaginary parts (the proofs of the other parts of this theorem are analogous).

In view of Theorem 12.1 it may be assumed that A and H are given by the right hand sides of (12.1) and of (12.2), respectively. From each block represented by a pair $(J_{\ell_j}(\alpha_j), \eta_j F_{\ell_j})$, $(j = 1, \dots, q)$ select the first $\ell_j/2$ unit coordinate vectors if ℓ_j is even, and the first $(\ell_j + \eta_j)/2$ unit coordinate vectors if ℓ_j is odd. Also select from each block

$$\left(\begin{bmatrix} J_{m_k}(\beta_k) & 0 \\ 0 & J_{m_k}(\bar{\beta}_k) \end{bmatrix}, F_{2m_k} \right), \quad k = 1, \dots, s,$$

the first m_k unit coordinate vectors. Denote the selected vectors (augmented by zero blocks) by $v_1, \dots, v_p \in \mathbb{C}^n$. Clearly, v_1, \dots, v_p are linearly independent, the subspace spanned by v_1, \dots, v_p is A -invariant, and Lemma 4.2 guarantees that $p = \text{In}_+(H)$. Moreover, $[v_i, v_k] = 0$ for all pairs $i, k = 1, \dots, p$, unless $i = k$ and v_i corresponds to the $(\ell_j + \eta_j)/2$ th unit coordinate vector in a block $(J_{\ell_j}(\alpha_j), \eta_j F_{\ell_j})$ with odd ℓ_j , in which case $[v_i, v_i] = 1$. It follows that $[x, x] \geq 0$ for every $x \in \text{Span}\{v_1, \dots, v_p\}$. Thus, the subspace $\mathcal{M}_+ := \text{Span}\{v_1, \dots, v_p\}$ satisfies all the requirements of Theorem 12.3. \square

For an invertible hermitian $H \in \mathbb{C}^{n \times n}$, a subspace $\mathcal{N} \subseteq \mathbb{C}^n$ is called H -nonnegative (resp., H -nonpositive) if $[x, x] \geq 0$ (resp., $[x, x] \leq 0$) for every $x \in \mathcal{N}$. The maximal dimension of an H -nonnegative (resp., H -nonpositive) subspace is equal to $\text{In}_+(H)$ (resp., $\text{In}_-(H)$). This fact can be proved by using the interlacing inequalities for eigenvalues of hermitian matrices; the details are omitted and the reader is referred to Chapter I.1 of [16], for example. Thus, the subspaces \mathcal{M}_+ and \mathcal{M}_- of Theorem 12.3 are in fact *maximal* (by set theoretic inclusion) H -nonnegative and *maximal* H -nonpositive, respectively.

13 Conclusions

Beginning with careful statements of the Jordan forms for real square matrices under similarity, and of the Kronecker canonical forms for matrix pairs under strict equivalence, a comprehensive review and analysis has been made of canonical forms for hermitian matrix pairs under strict equivalence and under congruence transformations. Careful attention has been paid to the distinct results applying to complex hermitian pairs over \mathbb{C} and to real symmetric pairs over \mathbb{R} . Scattered contributions to the subject have been synthesized in this work.

The analysis has also provided new insights into problems and methods of proof concerning diagonalization of hermitian matrix pairs, and concerning the question of when hermitian matrix pairs admit a positive semidefinite linear combination. Important connections have also been made with matrix structures appearing in the analysis of matrices which are selfadjoint in an *indefinite* inner product.

14 Acknowledgements

Peter Lancaster acknowledges support and encouragement from N. J. Higham, of the University of Manchester. The authors are grateful to anonymous referees for constructive comments.

References

- [1] Azizov, T. Ya., and Iokhvidov, I. S., *Linear Operators in Spaces with an Indefinite Metric*, John Wiley and Sons, Ltd., Chichester, 1989. (Translation of Russian original 1986.)
- [2] Au-Yeung, Y.-H., *A theorem on a mapping from a sphere to the circle and the simultaneous diagonalization of two hermitian matrices*, Proc. Amer. Math. Soc. **20** (1969), 545–548.
- [3] Au-Yeung, Y.-H., *Some theorems on simultaneous diagonalization of two Hermitian bilinear functions*, Glasnik Mat. **6(26)** (1971), 3–8.
- [4] Au-Yeung, Y.-H., and Poon, Y. T., *A remark on the convexity and positive definiteness concerning Hermitian matrices*, Southeast Asian Bull. Math. **3** (1979), 85–92.
- [5] Benner, P., Byers, R., Mehrmann, V., and Xu, H., *Numerical computation of deflating subspaces of skew-Hamiltonian/Hamiltonian pencils*, SIAM J. Matrix Anal. Appl. **24** (2002), 165–190.

- [6] Brinkman, L., *On the field of values of a matrix*, Proc. of Amer. Math. Soc., **12** (1961), 61–66.
- [7] Cheung, W.-S., Li, C.-K., and Rodman, L., *Operators with numerical range in a closed subspace*, submitted for publication.
- [8] Dieudonné, J., *Sur la reduction canonique des couples de matrices*, Bulletin de la Soc. Math. de France, **74** (1946), 130–146.
- [9] Finkbeiner, D., *Introduction to Matrices and Linear Transformations*, W. H. Freeman, San Francisco, 1978.
- [10] Finsler, P., *Über das Vorkommen definitiver und semidefiniter Formen in Scharen quadratischer Formen*, Comm. Math. Helv. **9** (1937), 188–192.
- [11] Franklin, J. N., *Matrix theory*, Prentice-Hall, Englewood Cliffs, N.J. 1968.
- [12] Gantmacher, F. R., *The Theory of Matrices*, Vols. 1 and 2, Chelsea, New York, 1959.
- [13] Gantmacher, F. R., and Krein, M. G., *Oscillation matrices and kernels and small vibrations of mechanical systems*, Revised edition. AMS Chelsea Publishing, Providence, RI, 2002. (Translation based on the 1941 original.)
- [14] Gohberg, I., Lancaster, P. and Rodman, L., *Spectral analysis of selfadjoint matrix polynomials*, Annals of Math., **112** (1980), 33-71.
- [15] Gohberg, I., Lancaster, P. and Rodman, L., *Matrix Polynomials*, Academic Press, New York, 1982.
- [16] Gohberg, I., Lancaster, P. and Rodman, L., *Matrices and Indefinite Scalar Products*, Birkhäuser Verlag, 1983.
- [17] Gohberg, I., Lancaster, P. and Rodman, L., *Invariant Subspaces of Matrices with Applications*, John Wiley (Canadian Math. Soc. Monographs), 1986.
- [18] Hestenes, M. R., *Pairs of quadratic forms*, Linear Algebra and Appl. **1** (1968), 397–407.
- [19] Horn, R. A., and Johnson, C. R., *Matrix Analysis*, Cambridge University Press, 1985.
- [20] Hua, L. K., *On the theory of automorphic functions of a matrix variable, II - The classification of hypercircles under the symplectic group*, Amer. J. Math, **66** (1944), 531–563.
- [21] Ingraham, M. H., *The singular case of the equivalence of pairs of Hermitian matrices*, Bulletin of Amer. Math. Soc., **40** (1934).

- [22] Ingraham, M. H., and Wegner, K. W., *The equivalence of pairs of Hermitian matrices*, Transactions of the Amer. Math. Soc., **38** (1935), 145–162.
- [23] Iohvidov, I. S., Krein, M. G., and Langer, H., *Introduction to the Spectral Theory of Operators in Spaces with an Indefinite Metric*, Akademie-Verlag, Berlin, 1982.
- [24] Krein, M. G., *The basic propositions in the theory of λ -zones of stability of a canonical system of linear differential equations with periodic coefficients*, Operator Theory: Advances and Applications, **7**, Birkhäuser, Basel, 1988. (Translation of Russian original of 1955.)
- [25] Krein, M. G., *Introduction to the geometry of indefinite J -spaces and to the theory of operators on those spaces*, AMS Translations, **93** (1970), 103–176. (Translation of Russian original of 1965.)
- [26] Krein, M. G., and Langer, H., *On some mathematical principles in the linear theory of damped oscillations of continua, I, II*, Integral Equations and Operator Theory, **1** (1978), 539–566. (Translation of Russian original of 1955.)
- [27] Kronecker, L., *Collected Works*, Chelsea, 1968.
- [28] Kronecker, L., *Algebraische Reduktion der Scharen bilinear Formen*, S. -B. Akad. Berlin, (1890), 763–776.
- [29] Lancaster, P., and Rodman, L., *Minimal symmetric factorizations of symmetric real and complex rational matrix functions*, Linear Algebra Appl., **220** (1995), 249–282.
- [30] Lancaster, P., and Rodman, L., *Canonical forms of real skewsymmetric matrix pairs under strict equivalence and congruence*, in preparation.
- [31] Lancaster, P., and Tismenetsky, M., *The Theory of Matrices*, Second Edition, Academic Press, Orlando, 1985.
- [32] Lancaster, P., and Ye, Q., *Inverse spectral problems for linear and quadratic matrix pencils*, Linear Algebra Appl., **107** (1988), 293-309.
- [33] Lancaster, P., and Ye, Q., *Variational properties and Rayleigh quotient algorithms for symmetric matrix pencils*, Operator Theory, Advances and Applications, **40** (1989), 247-278.
- [34] Langer, H., *Factorization of operator pencils*, Acta Sci. Math. (Szeged), **38** (1976), 83–96.
- [35] Lin, W. W., Mehrmann, V., and Xu, H., *Canonical forms for Hamiltonian and symplectic matrices and pencils*, Linear Algebra and its Applications, **302/303** (1999), 469–533.

- [36] Mac Duffee, C. C., *The Theory of Matrices*, Berlin, 1933, Chelsea, New York, 1946.
- [37] Malcev, A. I., *Foundations of Linear Algebra*, W. H. Freeman, San Francisco, 1961. (Translation of Russian original of 1948.)
- [38] Marcus, M., *Pencils of real symmetric matrices and the numerical range*, Aequationes Math., **17** (1978), 91–103.
- [39] Mehl, C., Mehrmann, V., and Xu, H., *Canonical forms for doubly structured matrices and pencils*, Electronic Journal of Linear Algebra **7** (2000), 112–151.
- [40] Mehrmann, V., and Xu, H., *Structured Jordan canonical forms for structured matrices that are Hermitian, skew-Hermitian, or unitary with respect to an indefinite inner product*, Electronic Journal of Linear Algebra, **5** (1999), 67–103.
- [41] Muth, P., *Über reele Äquivalenz von Scharen reeler Quadratischer Formen*, Crelle's Journal, **128** (1905), 302–343.
- [42] Pesonen, E. *Über die Spektraldarstellung quadratischer Formen in linearen Räumen mit indefiniter Metrik*, Ann. Acad. Sci. Fenn. Ser. A. **227**, (1956).
- [43] Pontryagin, L. S., *Hermitian operators in spaces with indefinite metric*, Izvestiya Akad. Nauk SSSR, Ser. Matem. **8** (1944), 243–280.
- [44] Ran, A. C. M., *Minimal factorization of selfadjoint rational matrix functions*, Integral Equations Operator Theory **5** (1982), 850–869.
- [45] Ran, A. C. M., and Rodman, L., *Stability of invariant Lagrangian subspaces. I.*, Oper. Theory Adv. Appl., **32**, 181–218, Birkhäuser, 1988.
- [46] Ran, A. C. M., and Rodman, L. *Stable invariant Lagrangian subspaces: factorization of symmetric rational matrix functions and other applications*, Linear Algebra Appl. **137/138** (1990), 575–620.
- [47] Rao, C. R., and Mitra, S. K., *Generalized Inverse of Matrices and its Applications*, Wiley, New York, 1971.
- [48] Rayleigh, J. W. S., *The Theory of Sound*, Third Edition, Dover, New York, 1945. (First Edition, 1877).
- [49] Shilov, G. E., *An Introduction to the Theory of Linear Spaces*, Prentice-Hall, Englewood Cliffs, 1961.
- [50] Smith, L., *Linear Algebra*, Springer Verlag, New York, 1984.

- [51] Thompson, R. C., *The characteristic polynomial of a principal subpencil of a hermitian matrix pencil*, Linear Algebra and its Applications, **14** (1976), 135-177.
- [52] Thompson, R. C., *Pencils of complex and real symmetric and skew matrices*, Linear Algebra and its Applications, **147** (1991), 323-371.
- [53] Trott, G. R., *On the canonical form of a non-singular pencil of Hermitian matrices*, Amer. J. of Math., **56** (1934), 359-371.
- [54] Turnbull, H. W., *On the equivalence of pencils of Hermitian forms*, Proc. of the London Math. Society, **39** (1935), 232-248.
- [55] Turnbull, H. W., and Aitken, A. C., *An Introduction to the Theory of Canonical Matrices*, 3rd Edition, Dover, New York, 1961.
- [56] Uhlig, F., *A recurring theorem about pairs of quadratic forms and extensions: a survey*, Linear Algebra Appl., **25** (1979), 219-237.
- [57] Wedderburn, J. H. M., *Lectures on Matrices*, Dover, New York, 1964, (First edition, 1934).
- [58] Wegner, K. W., *The equivalence of pairs of Hermitian matrices*, Trans. of Amer. Math. Soc., **38** (1935), 145-162.
- [59] Williamson, J., *The equivalence of nonsingular pencils of Hermitian matrices in an arbitrary field*, Amer. J. of Math., **57** (1935), 475-490.
- [60] Williamson, J., *On the algebraic problem concerning the normal forms of linear dynamical systems*, Amer. J. of Math., **58** (1936), 141-163.
- [61] Williamson, J., *Note on the equivalence of nonsingular pencils of Hermitian matrices*, Bulletin of Amer. Math. Soc., **51** (1945), 894-897.
- [62] Weierstrass, K., *Zur Theorie der quadratischen und bilinearen Formen*, Monatsber. Akad. Wiss., Berlin (1868), 310-338.