On condensed forms for partially commuting matrices

Yu. A. Alpin
Faculty of Mathematics and Mechanics
Kazan State University
420008 Kazan
Russia

L. Elsner
Fakultät für Mathematik
Universität Bielefeld
Postfax 100131
33501 Bielefeld
FRG

Kh. D. Ikramov [1]
Faculty of Computational Mathematics
and Cybernetics
Moscow State University
119899 Moscow
Russia

Abstract

Two complex $n \times n$ matrices A and B are said to be partially commuting if A and B have a common eigenvector. We propose a condensed form for such matrices that can be obtained from A and B by a finite rational computation. The condensed form is a pair of block triangular matrices, with the sizes of the blocks being uniquely defined by the original matrices. We then show how to obtain additional zeros inside the diagonal blocks of a condensed form by using the generalized Lanczos procedure as given by Elsner and Ikramov. This procedure can also be considered as a finite rational process. We point out several applications of the constructions above. It turns out that for Laffey pairs of matrices, i.e., for matrices (A, B) such that rank $[A, B] = 1$, the condensed form is a pair of 2×2 block triangular matrices. Using this fact, we show an economical way to find a spanning set for the matrix algebra generated by Laffey matrices A and B. Another application concerns so-called k-selfadjoint matrices. We examine such matrices in the unitary space as well as in a Krein space of defect 1. As an addendum, we give a new description of the Shemesh subspace of matrices A and B. This is the maximal common invariant subspace of A and B, on which these matrices commute.

[1] Research supported by Sonderforschungsbereich 343 *Diskrete Strukturen in der Mathematik*

1 Introduction

There exist many variations of the familiar notion of a commuting pair of matrices: nearly commuting matrices, quasi-commuting matrices, and so on. We wish to extend this list by a new term.

Definition Two complex $n \times n$ matrices A and B are said to be *partially commuting* if A and B have a common eigenvector.

One good reason for coining this term is obvious: if x is a nonzero vector such that

$$Ax = \lambda x, \qquad Bx = \mu x,$$

then there exist nontrivial common invariant subspaces of A and B, on which these matrices commute. The span of the vector x above is one of such subspaces.

A criterion for two complex matrices to have a common eigenvector was given by D. Shemesh in [7].

Theorem 1 Let $A, B \in M_n(\mathbf{C})$. Then A and B are partially commuting if and only if the subspace $\mathcal{N}$ defined by the formula

$$\mathcal{N} = \bigcap_{k,\ell=1}^{n-1} \ker\, [A^k, B^\ell] \tag{1}$$

is nontrivial.

Here and henceforth, the symbol $M_n(\mathbf{C})$ stands for the space of complex $n \times n$ matrices, and $[F, G]$ denotes the commutator of F and G: $[F, G] = FG - GF$.

In fact, the proof of Theorem 1 in [7] contains more than the statement above reveals. We state this extra as a separate theorem.

Theorem 2 The subspace (1) is invariant with respect to both matrices A and B, and A and B commute on $\mathcal{N}$. Moreover, $\mathcal{N}$ contains any other subspace of $\mathbf{C}^n$ with these properties.

It is our intention in this paper to point out some implications of Theorem 2 relating to the following problem.

Problem For given partially commuting matrices A and B, find matrices R and S with as many zeros as possible that can be obtained from the original matrices by a simultaneous similarity transformation

$$A \to R = P^{-1}AP, \qquad B \to S = P^{-1}BP. \tag{2}$$

Moreover, it is required that the similarity matrix P could be found by a finite computation employing only arithmetical operations. (For brevity, expressions like "P admits a finite rational computation" or "P can be computed rationally" are used in the paper.)

To make the point clearer, consider matrices A and B that commute on the whole space $\mathbf{C}^n$. It is almost universally known that such matrices are reducible by the same similarity to (upper) triangular form. However, the similarity matrix for this transformation cannot be computed rationally; moreover, it cannot in general be found in radicals. Indeed, the first column of such matrix must be an eigenvector for A (and for B as well). In Section 4, we show what can be done to commuting matrices A and B by a rational computation.

In Section 2, we prove that partially commuting matrices A and B can be brought by a simultaneous similarity to (say, upper) block triangular matrices

$$R = \begin{pmatrix} R_{11} & R_{12} & \dots & R_{1,k-1} & R_{1,k} \\ & R_{22} & \dots & R_{2,k-1} & R_{2,k} \\ & & \dots & \dots & \dots \\ & & & R_{k-1,k-1} & R_{k-1,k} \\ & & & & R_{kk} \end{pmatrix}, \tag{3}$$

$$S = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1,k-1} & S_{1,k} \\ & S_{22} & \dots & S_{2,k-1} & S_{2,k} \\ & & \dots & \dots & \dots \\ & & & S_{k-1,k-1} & S_{k-1,k} \\ & & & & S_{kk} \end{pmatrix}. \tag{4}$$

The diagonal blocks $R_{11}, R_{22}, \dots, R_{k-1,k-1}$ commute with the corresponding blocks $S_{11}, S_{22}, \dots, S_{k-1,k-1}$. The number k and the orders of diagonal blocks in (3)–(4) are uniquely defined by the matrices A and B. Moreover, the similarity matrix that brings A and B to R and S, respectively, can be computed rationally.

We call matrices (3) and (4) the *first condensed form* of the pair (A, B). This form looks especially nice for matrices A and B such that

$$\text{rank } [A, B] = 1. \tag{5}$$

Since matrices with property (5) were considered in an important paper [3] by Laffey, we call such A and B a *Laffey pair* of matrices.

In Section 3, we show how to refine the first condensed form of (A, B) in case where for the noncommuting diagonal blocks A_{kk} and B_{kk} in (3)–(4), the transposed blocks A_{kk}^T and B_{kk}^T commute. The resulting form is called the *second condensed form* of the pair (A, B).

In Section 5, we discuss some applications of the ideas in the previous sections. First, we consider the matrix algebra generated by a Laffey pair (A, B), and show how a basis of this algebra consisting of polynomials of the lowest possible degrees in A and B can be constructed. Then we examine so-called k-*selfadjoint* matrices. It follows from the results in [5] that such a matrix can be considered as a special pair of commuting complex matrices. Moreover, a similar description holds (at least, for $k = 2$) if we replace the ordinary conjugate transpose operation in the definition of k-selfadjoint matrices by the conjugation with respect to a Krein space of defect 1 [4]. Hence, the procedures in Section 4 apply to these matrices.

In an addendum, we give another characterization of the Shemesh subspace $\mathcal{N}$, which might be of independent interest.

2 The first condensed form of (A, B)

We begin by pointing out that the Shemesh subspace $\mathcal{N}$ for partially commuting A and B can be computed rationally. This is obvious from the characterization of $\mathcal{N}$ also found in [7]:

$$\mathcal{N} = \ker \sum_{k,\ell=1}^{n-1} [A^k, B^\ell]^* [A^k, B^\ell]. \tag{6}$$

Suppose that for given partially commuting matrices A and B, a basis $x_1, x_2, \ldots, x_{m_1}$ of $\mathcal{N}$ is already computed. If $m_1 = n$ then A and B commute on the whole space $\mathbf{C}^n$, and the first condensed form of (A, B) is these matrices themselves.

Assume that $m_1 < n$, and let P_1 be a nonsingular matrix with the first columns $x_1, x_2, \ldots, x_{m_1}$. The easiest way to construct such a matrix is to adjoin suitable coordinate vectors of the space $\mathbf{C}^n$ to the vectors $x_1, x_2, \ldots, x_{m_1}$. Consider the similarity transformation

$$A \to R_1 = P_1^{-1} A P_1, \qquad B \to S_1 = P_1^{-1} B P_1. \tag{7}$$

By the choice of P_1, the matrices R_1 and S_1 must have upper block triangular form,

$$R_1 = \begin{pmatrix} R_{11} & \tilde{R}_{12} \\ 0 & \tilde{R}_{22} \end{pmatrix}, \qquad S_1 = \begin{pmatrix} S_{11} & \tilde{S}_{12} \\ 0 & \tilde{S}_{22} \end{pmatrix}, \tag{8}$$

where the upper diagonal blocks R_{11} and S_{11} are of order m_1. Being the restrictions of the matrices A and B to the subspace $\mathcal{N}$, these blocks commute:

$$R_{11} S_{11} = S_{11} R_{11}.$$

It follows that in the corresponding block form of the commutator

$$C_1 = [R_1, S_1] = P_1^{-1} [A, B] P_1,$$

the upper diagonal block is zero:

$$C_1 = \begin{pmatrix} 0 & \tilde{C}_{12} \\ 0 & \tilde{C}_{22} \end{pmatrix}.$$

If $\tilde{R}_{22}$ and $\tilde{S}_{22}$ commute or do not *partially* commute, then matrices (8) are the first condensed form of (A, B). In the opposite case, a similarity transformation of type (7) can be applied to these blocks. It corresponds to a similarity transformation of the entire matrices R_1 and S_1:

$$R_2 = P_2^{-1} R_1 P_2, \qquad S_2 = P_2^{-1} S_1 P_2 \tag{9}$$

resulting in matrices of a refined block triangular form

$$R_2 = \begin{pmatrix} R_{11} & R_{12} & \tilde{R}_{13} \\ 0 & R_{22} & \tilde{R}_{23} \\ 0 & 0 & \tilde{R}_{33} \end{pmatrix}, \qquad S_2 = \begin{pmatrix} S_{11} & S_{12} & \tilde{S}_{13} \\ 0 & S_{22} & \tilde{S}_{23} \\ 0 & 0 & \tilde{S}_{33} \end{pmatrix}. \tag{10}$$

Again, the diagonal blocks R_{22} and S_{22} must commute:

$$R_{22}S_{22} = S_{22}R_{22}.$$

This implies that the commutator $C_2 = [R_2, S_2]$ is of the following form:

$$C_2 = \begin{pmatrix} 0 & C_{12} & \tilde{C}_{13} \\ 0 & 0 & \tilde{C}_{23} \\ 0 & 0 & \tilde{C}_{33} \end{pmatrix}. \tag{11}$$

Moreover, the block C_{12} *must be nonzero.* Assuming otherwise, we would quickly come to contradiction. Indeed, denoting by m_2 the order of the blocks R_{22} and S_{22}, and setting $m = m_1 + m_2$, we see that the span $\mathcal{E}_m$ of the first m coordinate vectors is a common invariant subspace of R_2 and S_2. Since $C_{12} = 0$, the matrices R_2 and S_2 commute on $\mathcal{E}_m$. This implies that the original matrices A and B have a common invariant subspace of dimension m, on which they commute. Moreover, this subspace properly contains $\mathcal{N}$, which contradicts Theorem 2.

Applying a suitable similarity to the blocks $\tilde{R}_{33}$ and $\tilde{S}_{33}$, and performing the corresponding transformation for the entire matrices R_2 and S_2, we obtain even more refined block triangular form than (10). Continuing in this way, we finally arrive at form (3)–(4) described in Introduction. Now we summarize the preceding discussion in a formal statement.

Theorem 3 Let $A, B \in M_n(\mathbf{C})$ be partially commuting matrices. Then for some positive integer k, A and B can be brought by a simultaneous similarity transformation (2) to block upper triangular form (3)–(4). In this form, the diagonal blocks $R_{11}, R_{22}, \ldots, R_{k-1,k-1}$ commute with the corresponding blocks $S_{11}, S_{22}, \ldots, S_{k-1,k-1}$. The last diagonal blocks R_{kk} and S_{kk} either commute or do not partially commute. The integer k and the orders of diagonal blocks are uniquely defined by the matrices A and B. Moreover, a similarity matrix P that brings A and B to R and S, respectively, can be computed rationally.

Remark We call k the *index* of the first condensed form of the pair (A, B).

Remark In fact, the argument in this section contains more than what is included in the statement of Theorem 3. Along with matrices (3)–(4), consider their commutator $T = [R, S]$. Then

$$T = \begin{pmatrix} 0 & T_{12} & \ldots & T_{1,k-1} & T_{1,k} \\ & 0 & \ldots & T_{2,k-1} & T_{2,k} \\ & & \ldots & \ldots & \ldots \\ & & & 0 & T_{k-1,k} \\ & & & & T_{kk} \end{pmatrix}, \tag{12}$$

and the off-diagonal blocks $T_{12}, T_{23}, \ldots, T_{k-2,k-1}$ *cannot be zero.* Also the blocks T_{kk} and $T_{k-1,k}$ cannot *both* be zero. This gives us the bound below.

Corollary Let A and B be partially commuting matrices, and let k be the index of their first condensed form. Then

$$\text{rank } [A, B] \geq k - 1. \tag{13}$$

As an illustration, we examine the case where A and B constitute a Laffey pair; i.e., relation (5) holds. We use an important result in [3].

Theorem 4 If $A, B \in M_n(\mathbf{C})$ satisfy (5) then A and B can be brought by a simultaneous similarity transformation to (upper) triangular form.

It follows from Theorem 4 and (5) that matrices A and B constituting a Laffey pair partially commute but do not commute. Hence, in Theorem 3 we have $k \geq 2$. By (13), $k \leq \text{rank } [A, B] + 1 = 2$, so $k = 2$. Now in the commutator for the first condensed form,

$$T = \begin{pmatrix} 0 & T_{12} \\ 0 & T_{22} \end{pmatrix}$$

the second block column is of rank 1. If $T_{22} \neq 0$ then rank $T_{22} = 1$, and R_{22} and S_{22} are a Laffey pair. But then the construction in Theorem 3 could be continued, leading to $k > 2$. Hence, $T_{22} = 0$, and the blocks R_{22} and S_{22} commute.

We summarize our findings in the theorem below.

Theorem 5 Let $A, B \in M_n(\mathbf{C})$ be a Laffey pair. Then the first condensed form of (A, B) are the matrices

$$R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix}, \tag{14}$$

where R_{ii} and S_{ii} commute for $i = 1, 2$. The commutator $T = [R, S]$ is a nilpotent matrix of the form

$$T = \begin{pmatrix} 0 & T_{12} \\ 0 & 0 \end{pmatrix}, \tag{15}$$

where the block T_{12} is of rank 1.

Remark If the rank of the commutator $[A, B]$ is interpreted as a measure of commutativity (or noncommutativity) of A and B, then Laffey pairs of matrices can be characterized as the closest ones to being commutative. Our Theorem 5 supports this characterization from a different point of view; namely, Laffey pairs of matrices have the least index of the first condensed form among all noncommutative matrix pairs.

3 Other condensed forms

The first condensed form of a pair (A, B) was defined as a pair of block upper triangular matrices. Of course, block lower triangular matrices could be used instead. To arrive at such condensed form, it suffices to change the top-to-bottom direction of reduction in Theorem 3 to the opposite direction. Then the following assertion is obtained.

Theorem 6 Let $A, B \in M_n(\mathbf{C})$ be partially commuting matrices. Then for some positive integer l, A and B can be brought by a simultaneous similarity transformation (2) to block lower triangular form

$$R = \begin{pmatrix} R_{11} & & & \\ R_{21} & R_{22} & & \\ \dots & \dots & \dots & \\ R_{l1} & R_{l2} & \dots & R_{ll} \end{pmatrix}, \qquad S = \begin{pmatrix} S_{11} & & & \\ S_{21} & S_{22} & & \\ \dots & \dots & \dots & \\ S_{l1} & S_{l2} & \dots & S_{ll} \end{pmatrix}. \qquad (16)$$

In matrices (16), the diagonal blocks $R_{22}, \dots, R_{ll}$ commute with the corresponding blocks $S_{22}, \dots, S_{ll}$. The first diagonal blocks R_{11} and S_{11} either commute or do not partially commute. The integer l and the orders of diagonal blocks are uniquely defined by the matrices A and B. Moreover, a similarity matrix that brings A and B to R and S, respectively, can be computed rationally.

Again, the theorem can be supplemented by asserting that the blocks $T_{32}, \dots, T_{l,l-1}$ of the commutator $T = [R, S]$ must be nonzero.

Now consider matrices A and B that may be not partially commuting but whose transposed matrices A^T and B^T do partially commute. Then another condensed form can be derived for the pair (A, B).

Theorem 7 Assume that for $A, B \in M_n(\mathbf{C})$ the transposed matrices A^T and B^T partially commute. Then for some positive integer m, A and B can be brought by a simultaneous similarity transformation (2) to block upper triangular form

$$R = \begin{pmatrix} R_{11} & R_{12} & \dots & R_{1,m-1} & R_{1,m} \\ & R_{22} & \dots & R_{2,m-1} & R_{2,m} \\ & & \dots & \dots & \dots \\ & & & R_{m-1,m-1} & R_{m-1,m} \\ & & & & R_{mm} \end{pmatrix}, \qquad (17)$$

$$S = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1,m-1} & S_{1,m} \\ & S_{22} & \dots & S_{2,m-1} & S_{2,m} \\ & & \dots & \dots & \dots \\ & & & S_{m-1,m-1} & S_{m-1,m} \\ & & & & S_{mm} \end{pmatrix}. \qquad (18)$$

In matrices (17)–(18), the diagonal blocks $R_{22}, \dots, R_{mm}$ commute with the corresponding blocks $S_{22}, \dots, S_{mm}$. The first diagonal blocks R_{11} and S_{11} either commute or do not partially commute. The integer m and the orders of diagonal blocks are uniquely defined by the matrices A and B. Moreover, a similarity matrix that brings A and B to R and S, respectively, can be computed rationally.

Proof Apply Theorem 6 to the partially commuting matrices A^T and B^T. Let the similarity transformation

$$K = Q^{-1}A^TQ, \qquad L = Q^{-1}B^TQ$$

give the block lower triangular matrices described in this theorem, with m being their index. Then the matrices

$$R = K^T = Q^TAQ^{-T}, \qquad S = L^T = Q^TBQ^{-T}$$

are the required condensed form for (A, B).

Remark Setting $T = [R, S]$ in Theorem 7, we can assert that the blocks $T_{23}, \ldots, T_{m-1,m}$ are nonzero. This implies a lower bound for the rank of $[A, B]$ similar to (13).

Returning to the first condensed form (3)–(4) of partially commuting matrices A and B, assume that the last diagonal blocks R_{kk} and S_{kk} do not commute, and the transposed blocks R_{kk}^T and S_{kk}^T partially commute. Then Theorem 7 can be applied to the pair (R_{kk}, S_{kk}) with a resulting refinement of matrices (3)–(4). This new pair of block upper triangular matrices will be called the *second condensed form* of the pair (A, B).

Remark The index of the second condensed form can be used to strengthen bound (13).

4 Further refinement of condensed forms

Let (R, S) be the first (or second) condensed form of the pair (A, B). Denote by k and $m_1, \ldots, m_k$ the index and the orders of diagonal blocks in the form. Consider a nonsingular block diagonal matrix

$$D = D_1 \oplus D_2 \oplus \cdots \oplus D_k, \tag{19}$$

where

$$D_i \in M_{m_i}(\mathbf{C}), \qquad i = 1, 2, \ldots, k.$$

Then, obviously, the matrices

$$\tilde{R} = D^{-1}RD, \qquad \tilde{S} = D^{-1}SD \tag{20}$$

give another condensed form for (A, B).

In this section, we explore the possibility to gain more zeros in the subdiagonal part of matrices R and S by using a freedom in choosing transformation (19)–(20). Recall that the diagonal blocks $R_{11}, \ldots, R_{k-1,k-1}$ commute with the corresponding blocks $S_{11}, \ldots, S_{k-1,k-1}$ if (R, S) is the first (upper) condensed form, with similar commutativity relations for other forms. Since the diagonal blocks are processed independently in the similarity transformation (19)–(20), we can state our problem as follows: for commuting $n \times n$ matrices A and B, find a simultaneous similarity transformation

$$A \to F = Q^{-1}AQ, \qquad B \to G = Q^{-1}BQ \tag{21}$$

that, first, produces as many zeros as possible in the subdiagonal part of both matrices F and G, and, second, can be computed rationally.

Below, we outline an approach to solving this problem based on the use of the generalized Lanczos procedure as given in [2]. Our exposition of this technique essentially parallels that in [2] differing in only one respect; namely, we admit any commuting matrices A and B instead of requiring that $B = A^*$.

Let $v \in \mathbf{C}^n$ be a fixed nonzero vector. Consider the vector sequence

$$v, Av, Bv, A^2v, ABv, B^2v, A^3v, A^2Bv, AB^2v, B^3v, A^4v, \ldots. \tag{22}$$

It is convenient to consider the sequence (22) as consisting of segments of length $1, 2, 3, 4, \ldots$, respectively. The $(k+1)$th segment will be called the *kth layer* of (22). It can be described as the totality of vectors of the form $u = W_k(A, B)v$, where $W_k(s, t)$ is any word of degree k in two (commuting) variables s and t; for $k = 0$, we set $W_0(A, B) = I_n$.

The subspace

$$L_m(A, B; v) = \text{span } \{W(A, B)v : \text{degree } (W) \leq m\} \tag{23}$$

is called the *mth generalized Krylov space*. We denote its dimension by l_m, and call $w_m = l_m - l_{m-1}$ the *width* of the mth layer. Here we set formally $w_0 = 1$. It is obvious that

$$w_m \leq m + 1. \tag{24}$$

Since $\{L_m(A, B; v)\}$ is a nested sequence of subspaces; i.e.,

$$L_0 \subset L_1 \subset L_2 \subset \ldots, \tag{25}$$

the numbers l_m constitute a nondecreasing sequence:

$$l_0 = 1 \leq l_1 \leq l_2 \leq \ldots. \tag{26}$$

Moreover, the first equality in (25) or (26) implies that from this point on, the sequences stabilize. For example, from the relations

$$1 < l_1 < l_2 < \ldots < l_p = l_{p+1} \tag{27}$$

one can deduce that

$$l_p = l_{p+1} = l_{p+2} = \cdots$$

and

$$L_p = L_{p+1} = L_{p+2} = \cdots .$$

If $l_p < n$ in (27) then L_p is a nontrivial common invariant subspace of A and B. Encountering such a subspace would be a lucky occasion; however, it is little probable. Thus, in what follows we assume that

$$l_p = n \tag{28}$$

in (27).

Definition The vectors

$$x_1, x_2, \ldots, x_n \tag{29}$$

are said to be a *graded basis* of $\mathbf{C}^n$ if for $q = 0, 1, \ldots, p$, the first l_q vectors in (29) form a basis of the generalized Krylov subspace L_q.

Consider A and B as linear operators on $\mathbf{C}^n$, and let (29) be a graded basis of $\mathbf{C}^n$. Then for both operators, their matrices with respect to (29) have a certain amount of zeros below the main diagonal. More precisely, for $i = 1, 2, \ldots$, the positions of zeros in the ith column can be determined by this rule: find the minimal index s of the subspaces L_q to which the vector x_i belongs, and set

$$N_i = i - 1 + w_s + w_{s+1}. \tag{30}$$

Then the entries with row indices exceeding N_i are zero.

It follows from (30) that among the subdiagonal entries of the ith column, only M_i entries, where

$$M_i = w_s + w_{s+1} - 1, \tag{31}$$

can be nonzero. This does not sound too optimistic, because w_s can grow linearly in accordance with (24). However, in some important particular cases, these numbers stabilize, beginning from some $s = s_0$.

Theorem 8 Assume that there exists an integer s_0 such that

$$w_s = s + 1, \qquad s < s_0,$$

$$w_{s_0} < s_0 + 1.$$

Then

$$w_s \le w_{s_0}, \qquad s > s_0.$$

The theorem is proved by repeating word-for-word the proof of Theorem 1 in [2], the only distinction being that the symbol A^* is replaced by B.

We give two simple examples where the stabilization described in Theorem 8 does occur.

Example 1 Each of the matrices A and B is a linear polynomial of the other. Then A and B can be simultaneously brought to Hessenberg form.

This statement is, of course, trivial: it suffices to bring, say, A to Hessenberg form, and under the same similarity, B assumes Hessenberg form as well. It is well known that the reduction of a matrix to Hessenberg form can be done rationally.

Nevertheless, it is interesting to look at this fact from the general point of view expressed by Theorem 8. Since

$$B = \alpha A + \beta I_n,$$

we have a linear relation in (22):

$$Bv = \alpha Av + \beta v.$$

This implies that Theorem 8 holds for (22) with $s_0 = 1$. Thus,

$$w_s \leq 1 \quad \text{for all } s,$$

and (see (31))

$$M_i \leq 1 \quad \text{for all } i.$$

Example 2 Suppose that (commuting) matrices A and B satisfy the quadratic relation

$$A^2 + B^2 = I_n.$$

Then for any vector v in (22), we have

$$w_2 \leq 2 < 3.$$

Hence,

$$w_s \leq 2, \qquad s > 2$$

and

$$M_i \leq 3 \quad \text{for all } i.$$

Definition A matrix $A \in M_n(\mathbf{C})$ is called *upper k-Hessenberg* if

$$a_{ij} = 0 \quad \text{for} \quad i > j + k.$$

Thus, the matrices of Example 2 can rationally be reduced to 3-Hessenberg matrices.

Summing up, we have shown a way to obtain additional zeros in the diagonal blocks of a condensed form of partially commuting matrices A and B. This is achieved by a finite rational computation, and the resulting form of the diagonal blocks sometimes resembles the Hessenberg form.

5 Some applications

5.1 Finding a spanning set for a Laffey algebra

For any two-generated subalgebra $\mathcal{A}(A,B)$ of $M_n(\mathbf{C})$, which is considered as a linear space, a spanning set can be found by a construction resembling the generalized Lanczos procedure in Section 4. Without loss of generality, we assume that $\mathcal{A}(A,B)$ contains the identity matrix I_n. Consider the matrix sequence

$$I_n, A, B, A^2, AB, BA, B^2, A^3, A^2B, BA^2, ABA, AB^2, BAB, B^2A, B^3, A^4, \dots \tag{32}$$

Each of these matrices can be written as

$$W_i(A,B), \tag{33}$$

where $W_i(s,t)$ is a word of degree i in two noncommuting variables s and t. The segment of sequence (32) formed by all words of degree k will be called the kth layer of (32). The subspace of $M_n(\mathbf{C})$ spanned by the layers $0,1,\dots,k$ is denoted by L_k, and the dimension of L_k is denoted by l_k.

Suppose that

$$L_0 \subset L_1 \subset L_2 \subset \cdots \subset L_p = L_{p+1}, \tag{34}$$

the symbol $\subset$ standing for the proper inclusion. Then

$$L_p = \mathcal{A}(A,B),$$

and the first $p+1$ layers of sequence (32) constitute a spanning set for the algebra $\mathcal{A}$.

It is obvious that

$$p \le n^2 - 1$$

whichever are the matrices A and B in (32). However, a better estimate is possible. It is shown in [6] that for general A and B,

$$p \le \lceil (n^2+2)/3 \rceil. \tag{35}$$

For some particular matrix pairs (A,B), much stronger bounds than (35) can be obtained. It is well known, for example, that

$$\dim \mathcal{A}(A,B) \le n \tag{36}$$

if $n \times n$ matrices A and B commute [1,8]. This implies that for such A and B,

$$p < n. \tag{37}$$

We are able to prove a bound comparable to (37) for another special class of matrix pairs; namely, for Laffey pairs of matrices.

Definition Let (A, B) be a Laffey pair of matrices. Then the two-generated algebra $\mathcal{A}(A, B)$ will be called a *Laffey algebra.*

Theorem 9 Let $\mathcal{A}(A, B)$ be a Laffey subalgebra of $M_n(\mathbf{C})$. Then for the number p in (34), the following bound holds:

$$p \leq 4n - 2. \tag{38}$$

Proof Without loss of generality, we can assume that A and B are in their first condensed form; that is,

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix}. \tag{39}$$

Then every matrix in $\mathcal{A}(A, B)$ has the same upper block triangular form.

Recall that for the commutator $C = [A, B]$, we have by Theorem 5

$$C = \begin{pmatrix} 0 & C_{12} \\ 0 & 0 \end{pmatrix}, \tag{40}$$

because the diagonal blocks A_{ii} and B_{ii} in (39) commute for $i = 1, 2$. It follows easily from (40) that

$$CXC = 0 \tag{41}$$

for every $X \in \mathcal{A}(A, B)$.

Consider any term

$$Z_1 Z_2 \cdots Z_t \tag{42}$$

in the t-th layer of (32). All factors Z_j in (42) assume only the values A and B. We claim that this product can be written as a sum of terms of the following two forms: $B^\alpha A^\beta, \alpha + \beta = t$, and $B^\alpha A^\beta C B^\gamma A^\delta, \alpha + \beta + \gamma + \delta = t - 2$, for some nonnegative integers α, β, γ, and δ. This is obvious for $t = 2$. In the induction step, we repeatedly use the fact that whenever a term AB occurs, it can be replaced by $BA + C$. By (41), all expressions containing two factors C vanish.

Now if $t = 4n - 1$, at least one of the exponents $\alpha, \beta, \gamma, \delta$ is not less than n. For such exponents, the Cayley-Hamilton theorem can be applied showing that in fact,

$$Z_1 Z_2 \cdots Z_t \in L_{4n-2}.$$

Thus, $L_{4n-1} = L_{4n-2}$, and $p \leq 4n - 2$.

Remark Note that in the proof above, we have not used the fact that rank $C_{12} = 1$ for a Laffey pair (A, B). Thus, the assertion of the theorem holds true for a broader class of matrix pairs than just Laffey matrices. These are pairs (A, B) reducible to 2×2 block triangular form, where the corresponding diagonal blocks commute.

5.2 k-selfadjoint matrices

The matrices in the title of this subsection are defined as follows:

Definition Let k be a positive integer. A matrix $A \in M_n(\mathbf{C})$ is called *k-selfadjoint* if

$$\sum_{i=0}^{k} (-1)^i \begin{pmatrix} k \\ i \end{pmatrix} A^{*i} A^{k-i} = 0. \tag{43}$$

Here, A^0 and A^{*0} are interpreted as the identity matrix I_n.

A complete description of k-selfadjoint matrices is given in [5].

Theorem 10 A matrix $A \in M_n(\mathbf{C})$ is k-selfadjoint if and only if

$$A = T + N, \tag{44}$$

where T is Hermitian,

$$N^{\lfloor \frac{k+1}{2} \rfloor} = 0, \tag{45}$$

and T and N commute.

Note that for 2-selfadjoint matrices, i.e., matrices A satisfying the relation

$$A^2 - 2A^*A + A^{*2} = 0, \tag{46}$$

(45) implies $N = 0$. Thus, 2-selfadjoint matrices are nothing else than ordinary Hermitian matrices. However, already for $k = 3$, relation (43) admits nonnormal solutions. Indeed, (45) assumes the form

$$N^2 = 0, \tag{47}$$

and the nilpotent part of A (see (44)) may be of index 2.

Let A be a 3-selfadjoint matrix. We apply the procedure in Section 4 to the commuting matrices T and N in (44). By virtue of (47), we should have

$$w_2 \le 2.$$

According to Theorem 8,

$$w_s \le 2, \qquad s > 2,$$

and

$$M_i \le 3 \qquad \text{for all } i.$$

It follows that T and N can simultaneously (and rationally!) be brought to 3-Hessenberg form. The same form is acquired by the 3-selfadjoint matrix A.

One can argue that, in fact, A can rationally be reduced to ordinary Hessenberg form (as can be any $n \times n$ matrix!). This is true; however, such a reduction does not generally give any special form to the matrices T and N in (44).

5.3 k-selfadjoint matrices in Krein spaces

Let $H \in M_n(\mathbf{C})$ be a Hermitian matrix such that

$$H^2 = I_n. \tag{48}$$

We exclude the cases $H = I_n$ and $H = -I_n$, and introduce an indefinite scalar product in $\mathbf{C}^n$ by the formula

$$[x, y] = (Hx, y) \qquad \forall x, y \in \mathbf{C}^n. \tag{49}$$

Here, $(\cdot, \cdot)$ denotes the ordinary scalar product in $\mathbf{C}^n$. The space $\mathbf{C}^n$ equipped with the scalar product (49) is called a *Krein space*.

For a matrix $A \in M_n(\mathbf{C})$ regarded as a linear operator on the Krein space $(\mathbf{C}^n; [\cdot, \cdot])$, the adjoint operator A^H is defined by the rule

$$[Ax, y] = [x, A^H y] \qquad \forall x, y \in \mathbf{C}^n.$$

We call A^H the *H-adjoint* matrix for A. A matrix A is *H-normal* if

$$AA^H = A^H A,$$

and *H-Hermitian* if

$$A = A^H.$$

Any matrix A in the Krein space $(\mathbf{C}^n; [\cdot, \cdot])$ can be written as

$$A = A_R + iA_J, \tag{50}$$

where A_R and A_J are H-Hermitian. Moreover, this representation of A is unique. Indeed,

$$A_R = \frac{1}{2}(A + A^H), \qquad A_J = \frac{1}{2i}(A - A^H).$$

A matrix A is H-normal if and only if the matrices A_R and A_J in (50) commute.

Replacing A^* by A^H in (43), we come to the definition of k-selfadjoint matrices in the Krein space $(\mathbf{C}^n; [\cdot, \cdot])$. In general, the properties of these matrices substantially differ from those of k-selfadjoint matrices in the unitary space. For example, 2-selfadjoint matrices need not now be H-Hermitian.

According to (48), the matrix H has only two eigenvalues: 1 and -1. Let q be the multiplicity of the eigenvalue -1. It is called the *defect* of the Krein space.

A description of 2-selfadjoint matrices in a Krein space was given in [4]. We reproduce it below as Theorem 11.

Theorem 11 Let A be a 2-selfadjoint matrix of a Krein space with defect $q > 0$; i.e.,

$$A^2 - 2A^H A + A^{H2} = 0.$$

Then the matrix A_J in decomposition (50) is a nilpotent matrix of rank at most $2q$, and

$$\text{rank }(A^H A - AA^H) \leq 2q - 1. \tag{51}$$

We point out some implications of this theorem and the preceding discussion. Assume that in Theorem 11,

$$q = 1.$$

Then (51) says that

$$\text{rank }(A^H A - AA^H) \leq 1.$$

In other words, the matrices A and A^H (or, if desired, A_R and A_J) either commute, which means that A is H-normal, or constitute a Laffey pair. In the latter case, the reduction in Section 2 can be applied to A and A^H. For the former case, consider the matrix sequence

$$I_n, A_R, A_J, A_R^2, A_R A_J, A_J^2, A_R^3, A_R^2 A_J, \ldots$$

According to Theorem 11,

$$A_J^3 = 0,$$

which implies (see Section 4) that

$$w_3 \leq 3.$$

Then

$$M_i \leq 5 \qquad \text{for all } i.$$

Hence, a matrix A that is H-normal and 2-selfadjoint can rationally be brought to 5-Hessenberg form, and the same form is acquired by the matrices A_R and A_J in (50).

Addendum

Let $A, B \in M_n(\mathbf{C})$ be partially commuting matrices, and $\mathcal{A}(A, B)$ the corresponding matrix subalgebra of $M_n(\mathbf{C})$. Again, we can assume that $I_n \in \mathcal{A}(A, B)$. Below, we give a description of the Shemesh subspace $\mathcal{N}$ that is different from the original definition (1).

Theorem 12 Let $C = [A, B]$ be the commutator of matrices A and B. Then the subspace $\mathcal{N}$ defined by (1) satisfies

$$\mathcal{N} = \bigcap_{Z \in \mathcal{A}(A,B)} \ker\,(CZ). \tag{52}$$

Proof Let $\mathcal{M}$ be the subspace in the right-hand side of (52). Then

$$x \in \mathcal{M} \quad \Longleftrightarrow \quad CZx = 0 \text{ for all } Z \in \mathcal{A}(A, B).$$

This makes the invariance of $\mathcal{M}$ with respect to every $Z \in \mathcal{A}(A, B)$ obvious. In particular, taking $Z = I_n$, we have

$$Cx = 0 \quad \text{for all } x \in \mathcal{M}.$$

Hence, A and B commute on $\mathcal{M}$.

Let $\mathcal{L}$ be another common invariant subspace of A and B such that these matrices commute on $\mathcal{L}$. Obviously, $\mathcal{L}$ is invariant with respect to every $Z \in \mathcal{A}(A, B)$. If y is a vector from $\mathcal{L}$, then $Zy \in \mathcal{L}$. Since A and B commute on $\mathcal{L}$, we have

$$AB(Zy) = BA(Zy),$$

or

$$C(Zy) = 0.$$

This argument applies to each $Z \in \mathcal{A}(A, B)$; hence, y must belong to $\mathcal{M}$. Since y is an arbitrary vector of $\mathcal{L}$, the inclusion

$$\mathcal{L} \subset \mathcal{M}$$

holds. Thus, $\mathcal{M}$ is the maximal common invariant subspace of A and B, on which these matrices commute. This is precisely the description of the Shemesh subspace $\mathcal{N}$.

Remark Contrary to (1), the formula (52) does not seem to be a finite description of $\mathcal{N}$. However, it can easily be made finite. It suffices to take the intersection in (52) over Z constituting a basis or a spanning set of the algebra $\mathcal{A}(A, B)$. It was shown in Subsection 5.1 that such a spanning set can be found be a finite rational computation.

References

1. J. Barria and P. R. Halmos, Vector bases for two commuting matrices, *Linear and Multilinear Algebra* 27:147–157 (1990).

2. L. Elsner and Kh. D. Ikramov, On a condensed form for normal matrices under finite sequences of elementary unitary similarities, *Linear Algebra Appl.* 254:79–98 (1997).

3. T. J. Laffey, Simultaneous triangularization of matrices — low rank cases and the nonderogatory case, *Linear and Multilinear Algebra* 6:269–306 (1978).

4. S. A. McCullough and L. Rodman, Two-selfadjoint operators in Krein spaces, *Integral Equations and Operator Theory* 26:202–209 (1996).

5. S. A. McCullough and L. Rodman, Hereditary classes of operators and matrices, *Amer. Math. Monthly* 104:415–430 (1997).

6. A. Paz, An application of the Cayley-Hamilton theorem to matrix polynomials in several variables, *Linear and Multilinear Algebra* 15:161–170 (1984).

7. D. Shemesh, Common eigenvectors of two matrices, *Linear Algebra Appl.* 62:11–18 (1984).

8. A. R. Wadsworth, The algebra generated by two commuting matrices, *Linear and Multilinear Algebra* 27:159–162 (1990).