On nonnormal matrices that can be reduced to band form under finite sequences of elementary unitary similarities

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

Kh. D. Ikramov*
Faculty of Computational Mathematics
and Cybernetics
Moscow State University
119899 Moscow
Russia

January 4, 1996

Abstract

We say that a square complex matrix is condensable if it can be reduced to a bandform through a finite sequence of elementary unitary similarities. The two main results of this paper are the following assertions:

(1) Any matrix with a low rank skew - Hermitian part is condensable.

(2) If A is a normal condensable matrix, then any rank one perturbation of A (generally nonnormal) is condensable as well.

*Research supported by Sonderforschungsbereich 343 *Diskrete Strukturen in der Mathematik*

1 Introduction

In this paper we continue the investigation of the generalized Lanczos procedure introduced in [Elsner, Ikramov]. In the latter, mostly the case of a normal matrix has been studied. It has been shown that the application of the generalized Lanczos procedure to a matrix A produces a variable bandwidth matrix B that we call the condensed form of A. If A is normal the bandwidth is still variable but considerably smaller. For some important particular cases (including, of course, Hermitian and unitary matrices), the condensed form is a genuine band matrix with a modest bandwidth. Moreover, there exist situations when this band matrix, rather broad in its upper left corner, shrinks ultimately to the bandwidth 3.

Not all normal matrices can be reduced to a constant bandwidth form through finite sequences of unitary similarities. This is even more so for general nonnormal matrices. In fact, even if *infinite* sequences of unitary similarities are admitted, the following assertion can be proved by a simple dimension argument using some basic facts from algebraic geometry: "Given a fixed bandwidth p, then for $n \geq$ some integer depending on p, there are complex $n \times n$ matrices which are not unitarily similar to a p-banded matrix" (here the short paper [Sturmfels] was cited).

Nevertheless, there exist nonnormal matrices that can be unitarily reduced even to tridiagonal form. Some classes of such matrices can be found in [Longstaff]. Motivated by these results, we describe in this report several classes of nonnormal matrices that can be transformed, essentially through our generalized Lanczos procedure, into constant bandwidth matrices. Although similar in spirit to the results in [Longstaff], ours are different in at least three respects: firstly, we do not confine ourselves to the bandwidth three; furthermore, we do insist our condensed forms to be achieved via *finite* sequences of elementary unitary similarities; and, finally, they can be constructed with an arbitrary choice of the initial vector.

The paper is organized as follows. In Section 2, we fill a gap in the previous paper. There two versions of our generalized Lanczos procedure had been presented. Here we show the relations between them, a topic treated there in a more informal manner. In Section 3 linear dependencies inside generalized Krylow sequences are studied. The main results are in Section 4. It is shown (Theorem 6 and 7) that any matrix with low rank skew - Hermitian part can be reduced to bandform through a finite sequence of elementary unitary similarities. A similar result holds for rank one perturbations of normal matrices (Theorem 10).

2 A generalized Lanczos procedure

In this section, we recall two of the three versions of our generalized Lanczos procedure. They are given in [Elsner, Ikramov] as Algorithm 1 and Algorithm 3, respectively. We explain the relation between these two algorithms (this was not done in [Elsner, Ikramov]).It turns out that Algorithm 3 is equivalent to Algorithm 1 for a particular ordering of the generalized Krylov vectors.

For a given $n \times n$ matrix A and a nonzero initial vector v, we are dealing with the vector sequence

$$v, Av, A^*v, A^2v, A^*Av, AA^*v, A^{*2}v, A^3v, \dots \tag{1}$$

called the generalized Krylov sequence. It can be divided into subsequences of length 1,2,4,8,16,..., respectively, corresponding to vectors of the form

$$u = W_k(A, A^*)v \tag{2}$$

where $W_k(s,t)$ is any word of degree k in two noncommuting variables s and t. The corresponding part of the sequence (1) is called the k-th layer.

Now and then, it will be more convenient to inspect the matrix sequence

$$I, A, A^*, A^2, A^*A, AA^*, A^{*2}, A^3, A^2A^*, AA^*A, \dots, A^{*3}, \dots \tag{3}$$

instead of the sequence (1). For the matrix sequence the layers are defined accordingly.

In the procedure below, we construct a sequence of orthogonal and nonzero vectors $v_1, v_2, v_3, \dots$ from the sequence (1). We set

$$\hat{\mathcal{K}}_m = \operatorname{span}\{v_1, \dots, v_m\}. \tag{4}$$

If Pw is the orthogonal projection of a vector w onto a subspace $\mathcal{K}$, then $w - Pw$ is the perpendicular from w on $\mathcal{K}$ and is denoted by

$$\operatorname{orth}_{\mathcal{K}} w.$$

Algorithm 1 (The generalized Lanczos procedure)

1. Choose a random nonzero vector v. Let $v_1 = v$.

2. Assume that orthogonal and nonzero vectors $v_1, \ldots, v_m$ have already been found, which constitute an orthogonal basis of the linear span of the first k_m vectors $u_1, \ldots, u_{k_m}$ of the sequence (1). If the perpendicular

$$\operatorname{orth}_{\hat{\mathcal{K}}_m} u_{k_m+1}$$

is nonzero, take it as the vector v_{m+1}. Otherwise, try

$$\operatorname{orth}_{\hat{\mathcal{K}}_m} u_{k_m+2},$$

and, if it is nonzero, take it as v_{m+1}, and so on.

Remark Our generalized Lanczos vectors $v_1, \ldots, v_m$ are, in general, not normalized. However, nothing prevents us from choosing a unit initial vector v and inserting normalization at any step of our procedure.

Remark If the perpendiculars from the vectors u_k of the current layer are all zero, then $\hat{\mathcal{K}}_m$ is a (common) invariant subspace of A and A^* and Algorithm 2 comes to an end.

To state Algorithm 3, some notation is needed; e.g. [Bunse–Gerstner, Elsner]. For any $A \in C^{n \times n}$ its columns and rows are denoted by the corresponding small letter as $a_{*1},\ a_{*2}, \ldots, a_{*n}$ and $a_{1*}, a_{2*}, \ldots, a_{n*}$, respectively. For $1 \leq j < n$ we denote by $P(j, z)$ the complex Householder transformation which eliminates the entries $j+1$ through n in the nonzero vector z and does not change the first $j-1$ entries of z.

Algorithm 3

1. Let $\ell_0 = 2,\ k = 1$. As long as $k \leq n$ do the following

2. Let $m = (k+1)/2$. Check the entries $\ell_{k-1}+1$ through n in the column $a_{*,m}$ of the current matrix A. If all these entries are zero let $\ell_k = \ell_{k-1}, k \to k+1$ and go to Step 3. Otherwise, find the Householder transformation $P_k = P(\ell_{k-1}, a_{*,m})$ and perform the similarity

$$A \to P_k A P_k. \tag{5}$$

   Let $\ell_k = \ell_{k-1} + 1$. If $\ell_k = n$ go to Step 4; otherwise, $k \to k+1$ and go to Step 3.

3. Let $m = k/2$. Check the entries $\ell_{k-1}+1$ through n in the row $a_{m,*}$ of the current matrix A. If all these entries are zero let $\ell_k = \ell_{k-1}, k \to k+1$ and go to Step 2. Otherwise, find the Householder transformation $P_k = P(\ell_{k-1}, a_{m,*})$ and perform the similarity (5). Let $\ell_k = \ell_{k-1}+1$. If $\ell_k = n$ go to Step 4; otherwise, $k \to k+1$ and go to Step 2.

4. Stop the procedure.

Generically, this algorithm will eliminate $n-2$ entries in the first stage, $n-3$ in the second stage, and so on, so that it stops after $n-2$ stages, when $(n-1)(n-2)/2$ zeros have been created in the matrix A. We note that this is the same amount of zeros as that in the standard reduction to Hessenberg form. The difference is that now we are placing zeros on both sides of the main diagonal of A.

Let B be the matrix at the end of Algorithm 3. If we let

$$P = P_1 P_2 \cdots P_N \tag{6}$$

where N is the overall number of stages in the algorithm, then

$$B = P^* A P. \tag{7}$$

We call B the condensed form of the matrix A.
Now we establish the relation between both algorithms under the simplifying assumption that all the pivot elements in B, i.e. entries

(2,1), (1,3), (4,2), (2,5), (6,3), (3,7), ...

are nonzero. We begin with this simple observation: if v is normalized and

$$\mathcal{K}_m(A, v)$$

denotes the linear span of the first m vectors of sequence (1), and $e_i (1 \leq i \leq n)$ are the coordinate vectors, then, since $Pe_1 = v$,

$$P\mathcal{K}_m(B, e_1) = \mathcal{K}_m(A, v). \tag{8}$$

Hence, we can obtain the full information of the generalized Krylov subspaces of the initial matrix A by inspecting the corresponding subspaces of its condensed form.

Let $\mathcal{L}_k$ be the linear span of the first k coordinate vectors e_i. Then our assumption on pivot elements is equivalent to the relations

$$\begin{aligned}
&\mathcal{L}_2 = \mathrm{span}\{e_1, Be_1\},\\
&\mathcal{L}_3 = \mathrm{span}\{e_1, Be_1, B^*e_1\},\\
&\mathcal{L}_4 = \mathrm{span}\{e_1, Be_1, B^*e_1, Be_2\},\\
&\mathcal{L}_5 = \mathrm{span}\{e_1, Be_1, B^*e_1, Be_2, B^*e_2\},\\
&\mathcal{L}_6 = \mathrm{span}\{e_1, Be_1, B^*e_1, Be_2, B^*e_2, Be_3\},
\end{aligned} \tag{9}$$

and so on. We show that these relations are, in turn, equivalent to the equalities

$$\begin{aligned}
&\mathcal{L}_2 = \mathrm{span}\{e_1, Be_1\} = \mathcal{K}_2(B, e_1),\\
&\mathcal{L}_3 = \mathrm{span}\{e_1, Be_1, B^*e_1\} = \mathcal{K}_3(B, e_1),\\
&\mathcal{L}_4 = \mathrm{span}\{e_1, Be_1, B^*e_1, B^2e_1\} = \mathcal{K}_4(B, e_1),\\
&\mathcal{L}_5 = \mathrm{span}\{e_1, Be_1, B^*e_1, B^2e_1, B^*Be_1\} = \mathcal{K}_5(B, e_1),\\
&\mathcal{L}_6 = \mathrm{span}\{e_1, Be_1, B^*e_1, B^2e_1, B^*Be_1, BB^*e_1\} = \mathcal{K}_6(B, e_1),\\
&\mathcal{L}_7 = \mathrm{span}\{e_1, Be_1, B^*e_1, B^2e_1, B^*Be_1, BB^*e_1, B^{*2}e_1\} = \mathcal{K}_7(B, e_1),
\end{aligned} \tag{10}$$

and so on.
The first two relations (10) need no proof. Since $e_2 \in \mathcal{L}_2$, we have

$$Be_2 \in B\mathcal{L}_2 = \mathrm{span}\{Be_1, B^2e_1\}.$$

Hence (see (9))

$$\mathcal{L}_4 \subset \mathrm{span}\{e_1, Be_1, B^*e_1, B^2e_1\},$$

and the subspace on the right-hand side must be four-dimensional and coincide with $\mathcal{L}_4$.
Similarly, from inclusions

$$\begin{aligned}
&B^*e_2 \in B^*\mathcal{L}_2 = \mathrm{span}\{B^*e_1, B^*Be_1\},\\
&Be_3 \in B\mathcal{L}_3 = \mathrm{span}\{Be_1, B^2e_1, BB^*e_1\},\\
&B^*e_3 \in B^*\mathcal{L}_3 = \mathrm{span}\{B^*e_1, B^*Be_1, B^{*2}e_1\}
\end{aligned} \tag{11}$$

we deduce

$$\begin{aligned}
&\mathcal{L}_5 \subset \mathrm{span}\{e_1, Be_1, B^*e_1, B^2e_1, B^*Be_1\},\\
&\mathcal{L}_6 \subset \mathrm{span}\{e_1, Be_1, B^*e_1, B^2e_1, B^*Be_1, BB^*e_1\},\\
&\mathcal{L}_7 \subset \mathrm{span}\{e_1, Be_1, B^*e_1, B^2e_1, B^*Be_1, BB^*e_1, B^{*2}e_1\},
\end{aligned}$$

with inclusions being in fact equalities.

We conclude that for any $k, 1 \le k \le n$ the first k coordinate vectors e_i constitute an orthogonal basis of the generalized Krylov subspace $\mathcal{K}_k(B, e_1)$. According to (8), this is equivalent to saying that the first k columns of the matrix P constitute an orthogonal basis of the generalized Krylov subspace $\mathcal{K}_k(A, v_1)$; here v_1 is the first column of P (in other words, v_1 is the initial vector of the generalized Lanczos procedure).

Remark The argument above can show clearly the distinction between a normal and nonnormal matrix A. Indeed, for A (and B) normal (11) gives

$$Be_3 \in B\mathcal{L}_3 = \mathrm{span}\{Be_1, B^2e_1, BB^*e_1\} = \mathrm{span}\{Be_1, B^2e_1, B^*Be_1\} \subset \mathcal{L}_5.$$

Hence, adjoining the vector Be_3 to $e_1, Be_1, B^*e_1, Be_2, B^*e_2$ does not increase the dimension of the linear span. This means that the entry b_{63} has to be zero.

Remark In our generalized Lanczos procedure, any ordering of vectors inside each layer of the sequence (1) can be admitted. This corresponds to the fact that, instead of Algorithm 3, the whole set of elimination algorithms could be described differing one from another by the order in which rows and columns of A are inspected. Consider, for example, the condensed form on the picture below.

$$\begin{array}{cccc}
\times & \times & \otimes_2 & \\
\otimes_1 & \times & \times & \otimes_3 \\
 & \times & \times & \cdots \\
 & \times & \times & \cdots \\
 & \otimes_4 & \times & \cdots \\
 & & \otimes_5 & \cdots \\
 & & & \cdots
\end{array}$$

Here the pivot elements are circled and indexed, according to the number of stage. We see that the first two stages of the corresponding elimination algorithm coincide with thoses of Algorithm 3. However, on Stage 3, we process the second row of the current matrix instead of the second column in Algorithm 3. The next two stages are eliminations on columns, and so on. Under the assumption above that all the pivot elements be nonzero, and essentially repeating the argument above, one can show that this modification of Algorithm 3 corresponds to the following ordering of generalized Krylov vectors:

$$e_1, Be_1, B^*e_1, B^*Be_1, B^{*2}e_1, B^2e_1, BB^*e_1, \ldots$$

3 Linear dependencies inside a layer

In the sequel, we will need some terminology introduced in [Elsner, Ikramov]. From this point on, the m-th generalized Krylov subspace is understood as the subspace

$$L_m(A, v) = \text{span}\{W(A, A^*)v : \text{degree}W \leq m\}. \tag{12}$$

We denote its dimension by l_m, and call $w_m = l_m - l_{m-1}$ the width of the m-th layer. We set formally $w_0 = 1$. These, or similar definitions we apply also to the sequence (3). It is clear that

$$w_m \leq 2^m \tag{13}$$

for both sequences (1) and (3).
In [Elsner, Ikramov], the following facts have been proved for a normal matrix A:

(a) If $N_i(M_i)$ denotes the number of nonzeros in the i-th column (i-th row) of the condensed form B of the matrix A, then

$$N_i \leq w_{s-1} + w_s + w_{s+1}, \qquad M_i \leq w_{s-1} + w_s + w_{s+1}. \tag{14}$$

Here s is the index of the layer to which the i-th generalized Lanczos vector v_i (see (4)) belongs.

(b) Let m_0 be the smallest among the indexes m of layers with the property that, in sequence (1), vectors of the layers 0 through m are linearly dependent. Then

$$w_s \leq w_{m_0} \qquad \text{for} \qquad s > m_0. \tag{15}$$

In the proof of (a) normality of the matrix is not used, so this fact holds for a general nonnormal matrix. For proving (b), however, normality is vital. (b) settles completely the question when the condensed form of a normal matrix converts into a constant width band.
Unfortunately, fact (b) does not hold for nonnormal matrices. We show this through a simple example below after some preliminary remarks.
If A is normal, and vectors (1) span the whole space C^n, then linear dependencies in the sequence (1) occur when and only when the corresponding dependencies (with the same coefficients) occur in the matrix sequence (3).

So, it is of no difference which of the two sequences to inspect. The situation for nonnormal matrices is more complicated. Of course, linear dependencies in (3) imply similar dependencies in (1), for any vector v. However, the converse statement is generally not true. This explains why, in the nonnormal case, we only discuss linear dependencies in (3). This discussion will be simplified by the following terminology also introduced in [Elsner, Ikramov]. The vectors $u_1,\ u_2, \ldots, u_s \in C^n$ are said to be linearly dependent over the linear subspace $L \subset C^n$ if there exist numbers $\alpha_1, \alpha_2, \ldots, \alpha_s$ not all zero, such that

$$\alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_s u_s \in L \tag{16}$$

which might be written also as

$$\alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_s u_s = 0 \ (\mathrm{mod}\ L). \tag{17}$$

If this relation is only possible with $\alpha_1 = \alpha_2 = \ldots = \alpha_s = 0$ then $u_1, u_2, \ldots, u_s$ are linearly independent over L.

Now assume that the matrix A generating (3) is normal. Normality is just a linear relation in the second layer of (3):

$$A^* A = A A^*, \tag{18}$$

and this reduces drastically the width of all subsequent layers so that instead of (13) we have

$$w_m \leq m + 1. \tag{19}$$

Still, if no other linear dependencies between members of (3) exist (assuming $l_m \leq n$), we have equality in (19), and the condensed form is not a constant bandwidth matrix.

Of course, there are well-known examples of linear relations with much more favourable implications for the condensed form. One can mention the Hermitian property ($A = A^*$), assuring a tridiagonal form, or unitarity ($A^* A = I$) which leads to a pentadiagonal matrix. We give another example, hopefully, not as overused as the ones above, and refering to a nonnormal matrix.

Suppose A is an oblique projector, i.e.

$$A^2 = A, \qquad A^* A \neq A A^*. \tag{20}$$

Starring the equality, we have

$$A^{*2} = A^*.$$

In other words,

$$A^2 = 0(\mathrm{mod} L_1), \qquad A^{*2} = 0(\mathrm{mod} L_1). \tag{21}$$

It follows that any new (say, the m-th) layer in (3) may give at most two new independent matrices, namely

$$(AA^*)^k \qquad \text{and} \qquad (A^*A)^k,$$

if $m = 2k$ and

$$A(A^*A)^k \qquad \text{and} \qquad (A^*A)^k A^*,$$

if $m = 2k + 1$. This means that the condensed form of A is a band matrix with the bandwidth at most 7.
A related example is an oblique idempotent, i.e. a matrix A such that

$$A^2 = I, \qquad A^*A \neq AA^*.$$

These matrices also satisfy both relations (21). In fact, for any projector A, the matrix $B = 2A - I$ is idempotent, and, for any idempotent B, the formula $A = (I + B)/2$ gives a projector A.
Remark One can ask if the relations like (20) give enough independent matrices to obtain a condensed form for the whole matrix A. In other words, can, for such a matrix, the sequence (1) span the whole C^n? It has been shown in [Doković] that the dimension of the algebra $\mathcal{A}(A, A^*)$ spanned by all monomials $A_1 A_2 \cdots A_k$ with each $A_i \in \{A, A^*\}$ equals $4d + \epsilon$, where d is the number of distinct eigenvalues of AA^* which are > 1, and ϵ is 1 or 0, according to whether 1 is an eigenvalue of AA^* or not. Therefore, if $d \geq n/4$ there are at least n independent matrices in (3). Using the canonical form for an oblique projector introduced in [Doković], one can show, again for $d \geq n/4$, that most choices of an initial vector v lead to sequences (1) spanning C^n.
For a normal matrix A, any linear relation in sequence (3), say,

$$\alpha_1 A^{s_1} A^{*t_1} + \alpha_2 A^{s_2} A^{*t_2} + \cdots + \alpha_l A^{s_l} A^{*t_l} = 0$$

implies the corresponding relation for any eigenvalue λ of A:

$$\alpha_1 \lambda^{s_1} \overline{\lambda}^{t_1} + \alpha_2 \lambda^{s_2} \overline{\lambda}^{t_2} + \cdots + \alpha_l \lambda^{s_l} \overline{\lambda}^{t_l} = 0.$$

For A nonnormal, the similar implication may not be true, if a monomial enters the linear relation with powers of A and A^*, interleaving each other more than once. However, note the following result by Kreyszig constituting a nonnormal generalization of the fact above:

Theorem 1 [Kreyszig] Let $f(x, y)$ be a polynomial in two variables x and y,

$$f(x,y) = \sum_{\mu=0}^{m} \sum_{\nu=0}^{l} a_{\mu\nu} x^{\mu} y^{\nu}. \tag{22}$$

Consider the equation

$$\sum_{\mu=0}^{m} \sum_{\nu=0}^{l} a_{\mu\nu} A^{\mu} A^{*\nu} = 0. \tag{23}$$

If A is a solution of (23), then, for any eigenvalue λ of A, we have

$$f(\lambda, \overline{\lambda}) = 0. \tag{24}$$

Proof If u is a left eigenvector of A, corresponding to λ, i.e.

$$u^T A = \lambda u^T,$$

then multiplying (23) by u^T on the left and by $\overline{u}$ on the right (with $\overline{u}$ meaning u conjugated elementwise) gives (24).

Remark A matrix A satisfying (23) needs not to be normal. For example, the equation

$$A^2 - A^{*2} = 0 \tag{25}$$

is satisfied by any oblique projector or idempotent matrix. Nevertheless, some relations for members of (3) do indeed imply normality of A. We mention two related results from [Hoffman, Taussky].

Theorem 2 Suppose that, for a complex $n \times n$ matrix A, complex numbers $a_1, a_2, b_1, b_2, c_1, c_2$ exist such that $c_1 + c_2 \neq 0$, and, with $\lambda_1, \ldots, \lambda_n$ being eigenvalues of A, the spectrum of the matrix

$$B = a_1 A + a_2 A^* + b_1 A^2 + b_2 A^{*2} + c_1 A^* A + c_2 A A^*$$

consists of the numbers

$$\mu_j = a_1 \lambda j + a_2 \overline{\lambda}_j + b_1 \lambda_j^2 + b_2 \overline{\lambda}_j^2 + (c_1 + c_2)|\lambda_j|^2, \qquad j = 1, \ldots, n.$$

Then A is normal.

Theorem 3 Under the conditions of Theorem 2, except that $c_1 + c_2 = 0$ is now possible, and with the relations

$$a_2 = \overline{a}_1, \quad b_2 = \overline{b}_1, \quad c_1, c_2 \in R, \quad c_1^2 + c_2^2 \neq 0$$

imposed on the numbers $a_1, \ldots, c_2$, the matrix A is normal.

We combine the above theorems to prove a sufficient criterion for normality.

Theorem 4 Suppose that

$$f(x, y) = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2,$$

and $a_{11} \neq 0$. Then any matrix A, satisfying the equation

$$f(A, A^*) = a_{00}I + a_{10}A + a_{01}A^* + a_{20}A^2 + a_{11}AA^* + a_{02}A^{*2} = 0$$

or

$$f(A^*, A) = a_{00}I + a_{10}A^* + a_{01}A + a_{20}A^{*2} + a_{11}A^*A + a_{02}A^2 = 0$$

is normal.

Proof Let A satisfy $B = f(A, A^*) = 0$. By Theorem 1 its eigenvalues λ_i satisfy $f(\lambda_i, \overline{\lambda_i}) = 0$. Hence for this B the assumptions of Theorem 2 are satisfied, so A is normal.
If $f(A^*, A) = 0$, then by the reasoning above A^* is normal. Hence A is normal.

Corollary Consider the plane quadric of the form

$$a_{11}\ x^2 + 2\ a_{12}\ xy + a_{22}\ y^2 = 1. \tag{26}$$

or, in the complex form,

$$cz^2 + \overline{cz}^2 + 2dz\overline{z} = 4 \tag{27}$$

where

$$c = a_{11} - a_{22} - (2a_{12})i, \quad d = a_{11} + a_{22}.$$

Define a class of normal matrices with eigenvalues belonging to the curve (26) by the relations

$$AA^* = A^*A,\ cA^2 + \overline{c}A^{*2} + 2dAA^* = 4I. \tag{28}$$

It follows from Theorem 4 that, if $d = a_{11} + a_{22} \neq 0$, we can omit the first relation in (28), and the second equation (28) alone gives us the same matrix class.

Remark The condition $a_{11} \neq 0$ in Theorem 4 cannot be omitted, as the equation (25) demonstrates.

Assume that, in the m-th layer of (3), we have a linear relation (over L_{m-1}) between monomials

$$A^{s_1}A^{*t_1}A^{s_2}A^{*t_2}\cdots A^{s_l}A^{*t_l}, \qquad s_1 + t_1 + \cdots + s_l + t_l = m. \tag{29}$$

We call the relation symmetric if, for any monomial (29) that enters this relation with a nonzero coefficient β, the monomial

$$A^{t_l}A^{*s_l}\cdots A^{t_1}A^{*s_1} \tag{30}$$

also enters, and for the corresponding coefficient γ the equality

$$|\beta| = |\gamma| \tag{31}$$

holds. Otherwise, the relation is called nonsymmetric. Equation (18) and the left equation in (21) are examples of a symmetric and nonsymmetric relation, respectively, in the second layer.

Theorem 5 For a nonsymmetric relation in the m-th layer, starring it gives a new relation linearly independent of the former.

Proof The assertion is obvious if there exists a pair of monomials (29) and (30), of which one enters the relation given, and the other not. Otherwise, the relation contains a pair of monomials (29) and (30) such that for the corresponding coefficients β and γ condition (31) does not hold. Then the new relation contains both these monomials with the coefficients $\overline{\gamma}$ and $\overline{\beta}$, respectively. Now the assertion required follows from the fact that

$$\begin{vmatrix} \beta & \gamma \\ \overline{\gamma} & \overline{\beta} \end{vmatrix} = |\beta|^2 - |\gamma|^2 \neq 0.$$

As an example, consider the relation

$$[AA^*A, A] = 0$$

or

$$A^2A^*A - AA^*A^2 = 0. \tag{32}$$

This equation gives one of the possible descriptions of square matrices A for which A^* and A^+ (the Moore-Penrose inverse of A) commute [Hartwig, Spindelboeck]. If A is nonsingular, than (32) is equivalent to the normality condition (18). A singular matrix A, satisfying (32) needs not to be normal. Since the relation (32) is obviously nonsymmetric, we can write another relation

$$A^*AA^{*2} - A^{*2}AA^* = 0$$

which is linearly independent with (32).

4 Some classes of nonnormal matrices that can be reduced to band form

We begin with the following important observation. Since

$$A = B + iC, \qquad A^* = B - iC \tag{33}$$

where B and C are Hermitian matrices

$$B = \frac{(A + A^*)}{2}, \qquad C = \frac{(A - A^*)}{2i}, \tag{34}$$

the same generalized Krylov subspaces will be obtained, if, instead of (3), we process the matrix sequence

$$I, B, C, B^2, BC, CB, C^2, B^3, B^2C, BCB, BC^2, CB^2, \ldots, C^3, \ldots \tag{35}$$

Theorem 6 Suppose that $\operatorname{rank}(C) = 1$ for a matrix A in (33). Then A can be unitarily reduced to band form with the bandwidth at most 7 by a finite process, employing only arithmetical operations and square roots.

Proof Let v be an arbitrary initial vector. Consider, instead of (1), the vector sequence

$$v, Bv, Cv, B^2v, BCv, CBv, C^2v, B^3v, B^2Cv, \ldots \tag{36}$$

The layers in this sequence are of width one, up to the moment when a nonzero vector first appears with the matrix C as the leftmost factor, and of width two, after that. If (36) spans C^n, then the condensed form of A is a block tridiagonal matrix with diagonal blocks of order (at most) two which implies the assertion of the theorem. In the opposite case, A will be decomposed into direct sum with the first diagonal block in band form. Then we repeat the procedure above for the second diagonal block, and so on.

Remark Let t be a nonzero vector from the range of A. Such a vector can be found trivially; for example, any nonzero column of A will do. Taking $v = t$, we observe that (36) is essentially (i.e. from the point of view of Krylov subspaces) equivalent to the ordinary Krylov sequence

$$t, Bt, B^2t, B^3t, \ldots$$

If the latter spans C^n, we get a tridiagonal form for our (non-Hermitian!) matrix A. That is one of results of the paper [Longstaff]. Otherwise, a tridiagonal form for the upper diagonal block will be obtained, with perhaps broader band form for the rest of the matrix.

Essentially the same line of reasoning as in Theorem 6 can be applied to prove the following more general result:

Theorem 7 Suppose that $\mathrm{rank}(C) = k \ll n$ for an $n \times n$ matrix A in (33). Then A can be unitarily reduced to band form with the bandwidth at most $4k+3$ by a finite process, employing only arithmetical operations and square roots.

Proof We remind the reader that w_i is the increase in dimension over L_{i-1} for the generalized Krylov subspace L_i. Denote by r_i and s_i the part of this increase due to vectors in the i-th layer of sequence (36) with the leftmost factor B and C, respectively. Obviously,

$$w_i = r_i + s_i. \tag{37}$$

If m is the maximum value of those indices i for which $l_i \neq 0$ then

$$\sum_{i=1}^{m} s_i \leq k. \tag{38}$$

The i-th layer in (36) can be obtained by applying first B and then C to the vectors of the $(i-1)$-th layer. This makes evident the relation

$$r_i \leq w_{i-1} = r_{i-1} + s_{i-1}. \tag{39}$$

Combining (37) through (39) gives

$$\begin{aligned} w_m = r_m + s_m \leq r_{m-1} + s_{m-1} + s_m \leq \ldots \leq \\ r_1 + s_1 + \cdots + s_m \leq r_1 + k \leq 1 + k. \end{aligned}$$

Since $s_i = 0$ for $i > m$, we have

$$r_i = w_i \leq w_{i-1} \leq \cdots \leq w_m \leq 1 + k, \qquad i > m.$$

We conclude that the condensed form of A is a block tridiagonal matrix with diagonal blocks of order (at most) $k+1$. With this the assertion of Theorem 7 is proved.

Now we consider a slightly more general situation than in Theorem 6, namely the case of a matrix of the form

$$A = N + R \tag{40}$$

with N normal, and R of rank 1. Note that R is not assumed to be normal, i.e. in the rank decomposition of R,

$$R = xy^*, \qquad x, y \in C^n \tag{41}$$

the vectors x and y are, in general, linear independent.
For an arbitrarily chosen vector v, define the subspaces M_i as follows:

$$M_1 = \text{span}\{x, y, v, Nv, N^*v\}, \tag{42}$$

$$M_{i+1} = \text{span}\{M_i, NM_i, N^*M_i\}, \qquad i = 1, 2, \ldots \tag{43}$$

Denote by $L_i, \tilde{L}_i$, and $\hat{L}_i$ the i-th generalized Krylov subspace generated by applying N to the vectors v, x, and y, respectively.

Lemma 8 $L_i \subset M_i \subset \text{span}\{L_i, \tilde{L}_{i-1}, \hat{L}_{i-1}\}$, $i = 1, 2, \ldots$

Proof For $i = 1$ the inclusions above are obvious, because of (42) (note that $L_1 = \text{span}\{v, Nv, N^*v\}, \tilde{L}_0 = \text{span}\{x\}, \hat{L}_0 = \text{span}\{y\}$). For greater values of i, the inclusions can be easily proved by induction. We only mention that formula (43) giving us the subspaces M_i mimics the way through which generalized Krylov subspaces are generated.
Consider now the matrix sequence

$$I, N, N^*, N^2, NN^*, N^{*2}, N^3, N^2N^*, NN^{*2}, N^{*3}, \ldots \tag{44}$$

with $w_i (i = 0, 1, 2, \ldots)$ denoting the width of the i-th layer. We define also numbers

$$\bar{l}_i = \text{dim} M_i, \qquad i = 1, 2, \ldots$$

$$\bar{w}_1 = \bar{l}_1, \qquad \bar{w}_i = \bar{l}_i - \bar{l}_{i-1}, \qquad i = 2, 3, \ldots$$

Lemma 9 $\bar{w}_i \leq w_i + 2w_{i-1}, \quad i = 2, 3, \ldots \quad \bar{w}_1 \leq 3 + w_1.$

Proof The assertion of this lemma is an immediate consequence of (42) and Lemma 8.

Theorem 10 Suppose that for a normal $n \times n$ matrix N the layers in sequence (44) are uniformly bounded by a constant b not depending on n:

$$w_i \leq b, \quad i = 1, 2, 3, \ldots$$

Then any (generally nonnormal) matrix A of the form (40) can be reduced to band form with the bandwidth at most $12b - 1$ by a finite process, employing only arithmetical operations and square roots.

Proof We apply the generalized Lanczos procedure to the sequence of subspaces $M_1, M_2, \ldots$ This will produce an orthogonal basis in which A is represented by a block tridiagonal matrix B with block sizes $\bar{w}_i$. This has been shown in ([Elsner, Ikramov], Ch. 2 and 3) for the case of the generalized Krylow subspaces, however it holds here as well. The main condition for this to hold is $AM_i \subset M_{i+1}$, $A^*M_i \subset M_{i+1}$, which is easily established using (43). By Lemma 9 we have $\bar{w}_i \leq 3b$, and hence B can be considered as a band matrix with bandwidth $12b - 1$.

We give a small example of a nonnormal matrix A satisfying the conditions of Theorem 10. The Jordan block

$$A \equiv J_n(0) = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & \cdots & & & \\ & & & \cdots & & \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}$$

is a rank-one correction of the cyclic permutation

$$N \equiv \mathcal{P}_n = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & \cdots & & & \\ & & & \cdots & & \\ & & & & 0 & 1 \\ 1 & & & & & 0 \end{bmatrix}$$

Notice that, while the algebra $\mathcal{A}(I, N, N^*)$ generated by I, N, and N^* is at most n-dimensional for any $n \times n$ normal matrix N, we can easily show that $\mathcal{A}(I, J_n, J_n^*)$ coincides with the full algebra $C^{n,n}$.
In connection with this example, we mention the recent result in [Krupnik]: if N is a normal matrix with simple eigenvalues, then N can be represented as a sum of a rank-one matrix and a nilpotent one.

References

1. L. Elsner and Kh.D. Ikramov *On a condensed form for normal matrices under finite sequences of elementary unitary similarities* Linear Algebra Appl. (submitted)

2. B. Sturmfels *Tridiagonalization of complex matrices and a problem of Longstaff* Linear Algebra Appl., Vol. 109, p.165–166 (1988)

3. W.E. Longstaff *On tridiagonalization of matrices* Linear Algebra Appl., Vol. 109, p.153–163 (1988)

4. A. Bunse–Gerstner and L. Elsner *Schur parameter pencils for the solution of the unitary eigenproblem* Linear Algebra Appl., Vol. 154–156, p.741–778 (1991)

5. D.Ž. Doković *Unitary similarity of projectors* Aequationes Mathematicae, Vol. 42, p.220–224 (1991)

6. E. Kreyszig *Ein Satz über Matrixeigenwerte* Elem. Math., Vol. 25(3), p.60–61 (1970)

7. A.J. Hoffman and O. Taussky *A characterization of normal matrices* J. Res. Nat. Bur. Standards, Vol. 52, p.17–19 (1954)

8. R.E. Hartwig and K. Spindelboeck *Matrices for which A^* and A^+ commute* Linear and Multilinear Algebra, Vol. 14, p.241–256 (1984)

9. M. Krupnik *Changing the spectrum of an operator by perturbation* Linear Algebra Appl., Vol. 167, p.113–118 (1992)