ORTHOGONAL HESSENBERG REDUCTION AND ORTHOGONAL KRYLOV SUBSPACE BASES[*]

JÖRG LIESEN[†] AND PAUL E. SAYLOR[‡]

Abstract. We study necessary and sufficient conditions that a nonsingular matrix A can be B-orthogonally reduced to upper Hessenberg form with small bandwidth. By this we mean the existence of a decomposition $AV = VH$, where H is upper Hessenberg with few nonzero bands, and the columns of V are orthogonal in an inner product generated by a hermitian positive definite matrix B. The classical example for such a decomposition is the matrix tridiagonalization performed by the hermitian Lanczos algorithm, also called the orthogonal reduction to tridiagonal form. Does there exist such a decomposition when A is nonhermitian? In this paper we completely answer this question. The related (but not equivalent) question of necessary and sufficient conditions on A for the existence of short-term recurrences for computing B-orthogonal Krylov subspace bases was completely answered by the fundamental theorem of Faber and Manteuffel [*SIAM J. Numer. Anal.*, 21 (1984), pp. 352–362]. We give a detailed analysis of B-normality, the central condition in both the Faber–Manteuffel theorem and our main theorem, and show how the two theorems are related. Our approach uses only elementary linear algebra tools. We thereby provide new insights into the principles behind Krylov subspace methods, that are not provided when more sophisticated tools are employed.

Key words. linear systems, Krylov subspace methods, Hessenberg reduction, matrix decomposition, short-term recurrences, normal matrices, B-normality

AMS subject classifications. 15A21, 15A23, 65F10, 65F25

DOI. 10.1137/S0036142903393372

1. Introduction. The decompositional approach to matrix computations, formalized by Householder in the 1950s, is counted among the "Top 10" algorithmic ideas of the 20th century [2]. One of the best known of these decompositions is the tridiagonalization of a nonsingular hermitian matrix A; see, e.g., [7, Chapter 9.1.2]. In a nutshell (and without specifying the respective matrix dimensions), for each nonzero vector v there exists a matrix V with first column v, and a square tridiagonal matrix H, such that $AV = VH$, and the columns of V are mutually orthogonal in the Euclidean inner product. This decomposition is computed by the hermitian Lanczos algorithm, and is sometimes called the *orthogonal reduction to tridiagonal form*. It is easy to see, by comparing columns in the matrix equation $AV = VH$, that the first j columns of V form a basis of $\mathcal{K}_j(A, v) \equiv \text{span}\{v, \ldots, A^{j-1}v\}$, the jth Krylov subspace

generated by A and v. The importance of this reduction from a theoretical as well as from a practical point of view can hardly be underestimated.

When A is nonhermitian, we are naturally led to ask for generalizations of the orthogonal reduction to tridiagonal form. Specifically, we ask if there exists a hermitian positive definite (HPD) matrix B such that a nonsingular nonhermitian A can still be B-orthogonally reduced to an upper Hessenberg matrix *with small bandwidth*. B-orthogonally here means that the columns of V are orthogonal in the B-inner product.

This paper studies necessary and sufficient conditions on A that guarantee the existence of such a B-orthogonal reduction. Our subject seems to be elementary, and one might suspect that it is covered in many textbooks on numerical linear algebra. However, while it appears to be common knowledge that the orthogonal reduction to tridiagonal form does not exist in general, see, e.g., [7, p. 499], we are not aware of any publication where the potential for generalizations has been thoroughly studied.

On the other hand, the related question of necessary and sufficient conditions on A for the existence of a short-term recurrence for computing B-orthogonal Krylov subspace basis vectors was completely solved by the fundamental theorem of Faber and Manteuffel [4]. Denoting the columns of V by v_j, we say that these vectors can be computed by an $(s + 2)$-term recurrence, when only the previous $s + 1$ vectors, $v_{j-s}, \ldots, v_j$, are required to compute v_{j+1}. For example, if the matrix H in the decomposition $AV = VH$ is tridiagonal, then the vectors v_j are computed by a 3-term recurrence. This is a key recurrence in many algorithms, including the famous conjugate gradient method. One immediately expects that for a given matrix A the existence of an $(s+2)$-term recurrence for computing a B-orthogonal Krylov subspace basis is equivalent to B-reducibility of A to upper Hessenberg form with bandwidth $s + 2$. However, due to intricate details that are easily overlooked, this expectation is in general false.

Our paper has the following goals. First, we give a thorough analysis of the B-reducibility of a nonsingular matrix A to upper Hessenberg form with small bandwidth. This is an interesting matrix property that apparently was not studied previously. Despite common belief, the necessary and sufficient conditions so that A has this property are not the same as the necessary and sufficient conditions in the Faber–Manteuffel theorem. This situation deserves to be clarified. Second, the proofs in this paper use standard tools of linear algebra only. We thereby hope to provide some additional insight into the necessity of the conditions in the Faber–Manteuffel theorem, for which no elementary (linear algebra based) proof is known. Third, we intend to improve the understanding of B-normality, the central necessary and sufficient condition in our context, by completely characterizing the set of HPD matrices B with respect to which a given matrix A is B-normal. Finally, our goal is to help in the general understanding of the foundations of and principles behind Krylov subspace methods.

The paper is organized as follows. In section 2 we discuss the basic algorithm for B-orthogonal Hessenberg reduction of a matrix and for computing B-orthogonal Krylov subspace bases. In section 3 we explain the sufficiency of B-normality in our context, and study this important concept in detail. In section 4 we discuss the necessity of B-normality. In section 5 we relate our results to the Faber–Manteuffel theorem and the existence of short-term recurrences for computing B-orthogonal Krylov subspace bases. Concluding remarks in section 6 close the paper.

Throughout the paper we assume exact arithmetic. In particular, the word

"computation" in this paper does not refer to a finite precision computation.

2. B-orthogonal reduction to upper Hessenberg form. Let A be any non-singular N by N matrix, let v_1 be any nonzero N-vector (v_1 is assumed to be nonzero to exclude trivialities), and let

$$
(2.1) \qquad \mathcal{K}_j(A, v_1) \equiv \operatorname{span}\{v_1, Av_1, \ldots, A^{j-1}v_1\} \quad \text{for} \quad j = 1, 2, \ldots,
$$

denote the jth Krylov subspace generated by A and v_1. It is well known that the $\mathcal{K}_j(A, v_1)$ form a nested sequence of subspaces of increasing dimension, and that there exists an index

$$
(2.2) \qquad d = d(A, v_1) \equiv \dim \mathcal{K}_N(A, v_1),
$$

which is often called the grade of v_1 with respect to A, for which

$$
\mathcal{K}_1(A, v_1) \subset \cdots \subset \mathcal{K}_{d-1}(A, v_1) \subset \mathcal{K}_d(A, v_1) = \mathcal{K}_{d+1}(A, v_1) = \cdots = \mathcal{K}_N(A, v_1).
$$

Furthermore, for each v_1, $d \leq d(A)$, where $d(A)$ denotes the degree of the minimal polynomial of A.

For any N by N HPD matrix B, the function $\langle \cdot, \cdot \rangle_B$, defined by $\langle x, y \rangle_B \equiv y^* B x$ for N-vectors x and y, is a positive definite inner product. Suppose that, for a given nonsingular matrix A, vector v_1, and HPD matrix B, we want to compute bases of the Krylov subspaces $\mathcal{K}_j(A, v_1)$, for $j = 1, 2, \ldots, d$, that are orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_B$ (B-orthogonal). In other words, we want to compute vectors $v_1, v_2, \ldots, v_d$ such that

$$
(2.3) \qquad \operatorname{span}\{v_1, \ldots, v_j\} = \mathcal{K}_j(A, v_1), \qquad j = 1, \ldots, d,
$$
$$
(2.4) \qquad \langle v_j, v_k \rangle_B = 0, \qquad j \neq k, j = 1, \ldots, d, k = 1, \ldots, d.
$$

Starting from v_1, this familiar and important task is performed by the following basic algorithm:

$$
(2.5) \qquad v_{k+1} = Av_k - \sum_{j=1}^{k} h_{jk} v_j, \qquad k = 1, \ldots, d,
$$

where

$$
(2.6) \qquad h_{jk} = \frac{\langle Av_k, v_j \rangle_B}{\langle v_j, v_j \rangle_B}.
$$

Apparently, this algorithm is nothing but the classical Gram–Schmidt implementation of Arnoldi's method; see, e.g., [7, Chapter 9.4.1].

Rewriting (2.5) in the form

$$
Av_k = v_{k+1} + \sum_{j=1}^{k} h_{jk} v_j, \qquad k = 1, \ldots, d,
$$

yields the matrix representation

$$
(2.7) \qquad A[v_1, \ldots, v_k] = [v_1, \ldots, v_{k+1}] \begin{bmatrix} h_{11} & \cdots & h_{1k} \\ 1 & \ddots & \vdots \\ & \ddots & h_{kk} \\ & & 1 \end{bmatrix},
$$

or

$$(2.8) \qquad AV_k = V_{k+1}H_{k+1,k}, \qquad k = 1, \ldots, d-1,$$

where $H_{k+1,k}$ is a $(k+1)$ by k unreduced upper Hessenberg matrix. Since $v_{d+1} = 0$, the matrix representation for $k = d$ may be written as

$$(2.9) \qquad AV_d = V_d \begin{bmatrix} h_{11} & \cdots & \cdots & h_{1d} \\ 1 & \ddots & & \vdots \\ & \ddots & \ddots & \\ & & 1 & h_{dd} \end{bmatrix} = V_d H_d,$$

where H_d is a d by d unreduced upper Hessenberg matrix. The B-orthogonality of the basis vectors, cf. (2.4), in this notation means that $V_d^* B V_d$ is a diagonal matrix.

For given A, v_1, and B, the decomposition (2.9) always exists and conditions (2.3) and (2.4) define it uniquely up to scaling of the columns of V_d. There are in fact several different algorithms that realize conditions (2.3) and (2.4). But in exact arithmetic all these algorithms lead to a decomposition of the form

$$A(V_d D) = (V_d D)(D^{-1}H_d D),$$

where V_d and H_d are as in (2.9), and D is a nonsingular diagonal matrix. Clearly, the nonzero pattern of H_d is invariant under diagonal similarity transformation, and thus, for given A, v_1, and B, conditions (2.3) and (2.4) lead to a uniquely defined nonzero pattern of H_d.

In the following we will be mostly interested in this pattern, particularly in the upper bandwidth of H_d. We call an upper Hessenberg matrix $(s+2)$-*band Hessenberg*, if it has no nonzero entries above its sth superdiagonal. (Here the 0th superdiagonal is the diagonal.) This gives rise to the following definition.

DEFINITION 2.1. *The nonsingular matrix A is B-reducible to $(s+2)$-band Hessenberg form if there exists an HPD matrix B such that for each v_1, either H_d in the decomposition* (2.9) *is $(s+2)$-band Hessenberg, or $d \leq s+1$.*

Note that for each nonsingular matrix A, $s \geq 0$ in Definition 2.1, since a 0- or 1-band Hessenberg matrix H_d is singular, which contradicts the nonsingularity of A. Hence B-reducibility to 2-band Hessenberg (lower bidiagonal) form is the one with smallest possible bandwidth that may occur. On the other hand, $s \leq d(A) - 1$ always holds as well, since no matrix H_d in (2.9) can possibly have more than $d(A) - 1$ superdiagonals. The condition $d \leq s+1$ in Definition 2.1 covers the trivial cases in which H_d has at most $s+2$ bands simply because it is of size at most $s+1$ by $s+1$.

The classical example is the one for $s = 1$, namely the reduction to tridiagonal form (tridiagonalization) with respect to the Euclidean inner product (i.e., $B = I$). For each nonsingular hermitian matrix A and each v_1 the decomposition (2.9) exists, where H_d is a tridiagonal (3-band Hessenberg) matrix and $V_d^* V_d$ is diagonal; see, e.g., [7, Chapter 9.1.2].

In the following sections we will study sufficient (section 3) and necessary (section 4) conditions that A is B-reducible to $(s+2)$-band Hessenberg form. We will then relate our result to the Faber–Manteuffel theorem, which gives necessary and sufficient conditions on A so that for each v_1, a B-orthogonal Krylov subspace basis can be computed by an $(s+2)$-term recurrence (section 5).

3. Sufficiency and characterization of B-normality. Let us consider the sufficient conditions that a given nonsingular matrix A is B-reducible to $(s+2)$-band Hessenberg form. A trivial sufficient condition is that the minimal polynomial of A has degree $d(A) \leq s + 1$. Then, for each v_1 we obtain $d \leq d(A) \leq s + 1$, and the second sufficient condition in Definition 2.1 is always satisfied. Thus, if $d(A) \leq s + 1$, then A is B-reducible to $(s+2)$-band Hessenberg form for each HPD matrix B.

If $d(A) > s + 1$, we require that there exists an HPD matrix B so that for each v_1 with $d > s + 1$ the matrix H_d in (2.9) is $(s+2)$-band Hessenberg, i.e., that

$$(3.1) \qquad h_{jk} = 0 \qquad \text{for} \quad j + s + 1 \leq k \leq d.$$

From (2.6) it follows that $h_{jk} = 0$ if and only if

$$0 = \langle Av_k, v_j \rangle_B = \langle v_k, A^+ v_j \rangle_B,$$

where $A^+ \equiv B^{-1} A^* B$ is the B-adjoint of A.

Now suppose that $A^+ = p_s(A)$ for a polynomial p_s of degree s. Then, since each v_j is of the form $v_j = p_{j-1}(A)v_1$,

$$A^+ v_j = p_s(A)p_{j-1}(A)v_1 \in \mathcal{K}_{j+s}(A, v_1).$$

But then B-orthogonality of v_k to $\mathrm{span}\{v_1, \dots, v_{k-1}\} = \mathcal{K}_{k-1}(A, v_1)$ for all $k \geq 2$ shows that (3.1) indeed holds for $j + s + 1 \leq k \leq d$. We formally state the nontrivial sufficient condition for B-reducibility of A to $(s+2)$-band Hessenberg form in the following definition.

DEFINITION 3.1. *If there exists an HPD matrix B such that*

$$A^+ \equiv B^{-1} A^* B = p_s(A)$$

for a polynomial p_s of degree s, then the matrix A is called normal of degree s with respect to B, or short B-normal(s).

Using this definition we can state our main theorem.

THEOREM 3.2. *The nonsingular matrix A is B-reducible to $(s+2)$-band Hessenberg form if and only if either A is B-normal(s), or $d(A) \leq s + 1$.*

Above we have shown sufficiency. Before we continue with necessity we will study the important concept of B-normality in more detail. We start with a collection of equivalent characterizations.

THEOREM 3.3. *For any matrix A the following are equivalent:*
(1) *There exists an HPD matrix B such that $A^+ = p(A)$ for a polynomial p.*
(2) *There exists an HPD matrix B such that $AA^+ = A^+A$.*
(3) *A is normalizable (similar to a normal matrix).*
(4) *A is diagonalizable.*
(5) *There exists an HPD matrix B such that A and A^+ (for this B) have the same complete set of B-orthogonal eigenvectors.*

Proof. (1) $\Rightarrow$ (2). Obviously, $p(A)A = Ap(A)$ for each polynomial p.

(2) $\Rightarrow$ (3). Assume (2) and define the matrix $\mathcal{A} \equiv B^{1/2}AB^{-1/2}$, to which A is similar. Then $\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}$, i.e., $\mathcal{A}$ is normal.

(3) $\Leftrightarrow$ (4). Suppose that A is normalizable, $A = S^{-1}MS$ with M normal. Since M is diagonalizable [8, Condition 11], A is diagonalizable as well. If A is diagonalizable, $A = WDW^{-1}$ with D diagonal, it is obviously similar to the normal matrix D.

(3) $\Rightarrow$ (1). Again assume that $A = S^{-1}MS$ with M normal. Define $B = S^*S$, which is HPD. By [8, Condition 17], $M^* = p(M)$ for some polynomial p, which implies $A^+ = (S^*S)^{-1}A^*(S^*S) = S^{-1}M^*S = S^{-1}p(M)S = p(A)$.

(3) $\Rightarrow$ (5). If $A = S^{-1}MS$ with M normal, define $B = S^*S$, which is HPD. By [8, Condition 11], there exist a unitary matrix U and a diagonal matrix D such that $M = U^*DU$. Hence, $A = (US)^{-1}D(US)$, where the columns of the matrix $(US)^{-1}$ form a complete set of eigenvectors of A. It is easy to see that these are B-orthogonal. In addition, A^+ for this B is of the form $A^+ = (US)^{-1}D^*(US)$, and hence has the same set of eigenvectors as A.

(5) $\Rightarrow$ (2). If there exists an HPD matrix B such that $A^+ = W\Lambda W^{-1}$ and $A = WDW^{-1}$, where Λ and D are diagonal, then $AA^+ = WD\Lambda W^{-1} = W\Lambda DW^{-1} = A^+A$. □

The implications in Theorem 3.3 have appeared in the literature before. In particular, the equivalence of (1), (2), (4), and (5) was proven in [5, Theorem 5 and Corollary 6], and later cited and used, for example, in [9, Theorem 4.4] and [1, p. 772]. However, the proofs in [5] are different from ours, as they do not directly make use of the fact that B-normality is equivalent to normalizability. Because we use this equivalence our proofs appear to be almost trivial, and the list of conditions in Theorem 3.3 can be easily extended by exploiting the lists of equivalent conditions of normality [3, 8], and rephrasing each such condition in terms of normalizability.

We will now characterize the HPD matrices B with respect to which a given (diagonalizable) matrix is normal. Clearly, the matrix B might not be uniquely defined. For example, if A itself is HPD, then it is normal with respect to $B = I$ and $B = A$.

As shown in Theorem 3.3, A is normal with respect to an HPD matrix B if and only if it has a complete set of B-orthogonal eigenvectors. Let these eigenvectors w_i be scaled to have B-norm one, i.e. $w_i^*Bw_i = 1$, and suppose the w_i are the columns of the matrix W. Then $W^*BW = \text{diag}\,(w_i^*Bw_i) = I$. This is equivalent to $B = (WW^*)^{-1}$. On the other hand, let $B = (WW^*)^{-1}$, where the columns of W form any complete set of eigenvectors of A. Then an easy calculation shows that $AA^+ = A^+A$, i.e., that A is normal with respect to B. We thus have proven the following theorem.

THEOREM 3.4. *Suppose that the matrix A is diagonalizable. Then the set of all HPD matrices B with respect to which A is normal is given by*

(3.2) $$\{\,(WW^*)^{-1}\;:\;W\;\;is\;an\;eigenvector\;matrix\;of\;A\,\}.$$

The characterization (3.2) allows us to derive an expression for the unique B-adjoint of A in case A is normal with respect to B.

COROLLARY 3.5. *Suppose that the matrix A is diagonalizable, $A = W\Lambda W^{-1}$, and that it is normal with respect to an HPD matrix B. Then the B-adjoint of A corresponding to this B is given by*

(3.3) $$A^+ = W\,\Lambda^*\,W^{-1}.$$

In particular, the B-adjoint of A is unique for all HPD matrices B with respect to which A is normal. Moreover, A is B-normal(s) if and only if A is diagonalizable and $p_s(\Lambda) = \Lambda^*$ for a polynomial of degree s.

Proof. Each HPD matrix B with respect to which A is normal is of the form (3.2). A direct computation shows that $A^+ = B^{-1}A^*B$ has the form (3.3), which is unique since it does not depend on the particular choice of W (similarly to A itself). For the second part of the corollary, suppose that A is B-normal(s), i.e., $A^+ = p_s(A)$

for a polynomial of degree s. Then A must be diagonalizable, $A = W\Lambda W^{-1}$, and (3.3) shows that $A^+ = W\Lambda^* W^{-1} = p_s(A) = Wp_s(\Lambda)W^{-1}$, which yields $p_s(\Lambda) = \Lambda^*$.

The proof of the reverse implication is similar. $\square$

Remark 3.6. The second part of this corollary was also derived in [9, Theorem B.1]. There the authors used a proof different from ours, and do not comment on the uniqueness of the B-adjoint in general.

As shown by Corollary 3.5, the B-normal degree of a diagonalizable matrix is determined by the location of its eigenvalues. It is well known that the B-normal(1) matrices are precisely the diagonalizable matrices that have all eigenvalues on a straight line in the complex plane [4, Lemma 3]. By sufficiency, each such matrix A is B-reducible to 3-band Hessenberg (tridiagonal) form. The standard examples are the hermitian and skew-hermitian matrices, that are all I-normal(1). Rare practical examples of B-normal(1) matrices that are normal with respect to an HPD matrix $B \neq I$ were derived in [6].

When the eigenvalues of the diagonalizable matrix A do not lie on a line, A must have B-normal degree $s > 1$. The question then arises about the lowest degree polynomial p_s for which $p_s(\Lambda) = \Lambda^*$. A recent result of Khavinson and Świątek [10] shows that each harmonic polynomial of the form $p_s(z) - \bar{z}$, where p_s is a polynomial of degree $s > 1$, has at most $3s-2$ complex zeros. Consequently, the class of B-normal(s) matrices with $s > 1$ contains diagonalizable matrices that may have at most $3s - 2$ distinct eigenvalues. This shows that the maximal size of the B-normal(s) matrices for small $s > 1$ is severely limited. We illustrate the results of this section by an example for $s = 3$.

Example 3.7. Consider the third degree harmonic polynomial

$$-\frac{1}{8}\,z(z^2 - 9) - \bar{z}, \quad \text{which has the 7 roots} \quad 0,\ \pm 1,\ \frac{\pm 5 \pm \sqrt{-7}}{2}\,.$$

We use the nonzero roots to define the diagonal nonsingular 6 by 6 matrix

$$A \;=\; \operatorname{diag}\left(\pm 1,\ \frac{1}{2}(\pm 5 \pm \sqrt{-7})\right).$$

By the second part of Corollary 3.5, this matrix is B-normal(3), and by (3.3) its unique B-adjoint is given by $A^+ = A^* = A$. Theorem 3.4 shows that A is normal with respect to all diagonal HPD matrices B. If we use any such matrix B and any v_1 with $d = 6$ in the basic algorithm (2.5)–(2.6), then sufficiency in Theorem 3.2 shows that the resulting matrix H_6 in (2.9) is 5-band Hessenberg. In fact, A is B-reducible to 5-band Hessenberg form for any diagonal HPD matrix B.

4. Necessary conditions. In this section we will show that the conditions in Theorem 3.2 are necessary. To avoid confusion, we will in this section denote the grade of the vector v_1 with respect to A by $d(A, v_1)$, cf. (2.2). Furthermore, we assume that the given nonsingular N by N matrix A is nonderogatory, i.e., that $d(A) = N$. This assumption is made for notational convenience, and without loss of generality. In case A is derogatory, we may in our derivation restrict to starting vectors v_1 with $d(A, v_1) = d(A)$ and all results will then hold for N replaced by $d(A)$. We start our discussion with proving an essential technical lemma.

LEMMA 4.1. *For a nonsingular and nonderogatory N by N matrix A there exists an HPD matrix B such that*

$$(4.1) \qquad A^+ v_1 \in \mathcal{K}_{s+1}(A, v_1) \qquad \textit{for all } v_1 \textit{ with} \quad d(A, v_1) = N\,,$$

if and only if either A is B-normal(s) for $0 \leq s \leq N - 2$, or $s = N - 1$.

Proof. We first prove sufficiency. Suppose that $s = N - 1$. Then for all v_1 with $d(A, v_1) = N$, $\mathcal{K}_{s+1}(A, v_1) = \mathcal{K}_N(A, v_1)$ is equal to the whole (real or complex) N-dimensional space and thus $A^+ v_1 \in \mathcal{K}_N(A, v_1)$ holds for all HPD matrices B. Now consider that $0 \leq s \leq N - 2$ and that A is B-normal(s). This means there exists an HPD matrix B for which $A^+ = p_s(A)$ for a polynomial p_s of degree s. Hence for each v_1, $A^+ v_1 = p_s(A)v_1 \in \mathcal{K}_{s+1}(A, v_1)$.

We next prove necessity, the harder part. Suppose that there exists an HPD matrix B such that (4.1) holds for some s with $0 \leq s \leq N - 2$ (in case $s = N - 1$ we are done). We need to show that A is B-normal(s).

Let $A = [X_1, \ldots, X_l] \operatorname{diag}(J_1, \ldots, J_l) [X_1, \ldots, X_l]^{-1}$ denote the Jordan canonical form of A, with the distinct eigenvalues $\lambda_1, \ldots, \lambda_l$. Since A is nonderogatory, only one Jordan block corresponds to each eigenvalue and, hence, each X_m, $m = 1, \ldots, l$, is a Jordan chain. Then it is easy to show that the vectors v that satisfy $d(A, v) = N$ are precisely the vectors that have a nonzero component corresponding to the last vector of *each* Jordan chain X_m. As a consequence, when we choose *any* v with $d(A, v) = N$, and *any* nonzero scalar γ with $\gamma \neq -\lambda_m$ for $m = 1, \ldots, l$, then the vector $w \equiv \gamma v + Av$ will satisfy $d(A, v) = d(A, Av) = d(A, w) = N$. In particular, we can find N linearly independent vectors v, such that v, Av, and $w \equiv \gamma v + Av$ for each nonzero $\gamma \neq -\lambda_m$, $m = 1, \ldots, l$, satisfy $d(A, v) = d(A, Av) = d(A, w) = N$.

Suppose that one such vector v is chosen, and let $w \equiv \gamma v + Av$ for some fixed nonzero $\gamma \neq -\lambda_m$, $m = 1, \ldots, l$. Then $d(A, w) = N$, so that by (4.1), $A^+ w \in \mathcal{K}_{s+1}(A, w)$. Hence there exist coefficients $\alpha_j^{(w)}$, $j = 0, \ldots, s$, such that

$$
\begin{aligned}
A^+ w &= \sum_{j=0}^{s} \alpha_j^{(w)} A^j w \\
&= \gamma \sum_{j=0}^{s} \alpha_j^{(w)} A^j v + \sum_{j=0}^{s} \alpha_j^{(w)} A^{j+1} v
\end{aligned}
$$

$$
(4.2) \qquad = \gamma \alpha_0^{(w)} v + \sum_{j=1}^{s} (\gamma \alpha_j^{(w)} + \alpha_{j-1}^{(w)}) A^j v + \alpha_s^{(w)} A^{s+1} v.
$$

Similarly, there exist coefficients $\alpha_j^{(v)}$ and $\alpha_j^{(Av)}$, $j = 0, \ldots, s$, such that

$$
\begin{aligned}
A^+ w &= \gamma A^+ v + A^+ (Av) \\
&= \gamma \sum_{j=0}^{s} \alpha_j^{(v)} A^j v + \sum_{j=0}^{s} \alpha_j^{(Av)} A^{j+1} v
\end{aligned}
$$

$$
(4.3) \qquad = \gamma \alpha_0^{(v)} v + \sum_{j=1}^{s} (\gamma \alpha_j^{(v)} + \alpha_{j-1}^{(Av)}) A^j v + \alpha_s^{(Av)} A^{s+1} v.
$$

Now note that since $d(A, v) = N$ and $0 \leq s \leq N - 2$, the vectors $v, \ldots, A^{s+1} v$ are linearly independent. Thus, the equality of (4.2) and (4.3) implies that

$$
(4.4) \qquad\qquad \alpha_0^{(v)} = \alpha_0^{(w)},
$$
$$
(4.5) \qquad \gamma \alpha_j^{(v)} + \alpha_{j-1}^{(Av)} = \gamma \alpha_j^{(w)} + \alpha_{j-1}^{(w)}, \qquad j = 1, \ldots, s,
$$
$$
(4.6) \qquad\qquad \alpha_s^{(Av)} = \alpha_s^{(w)}.
$$

Let us define $\eta_j \equiv \alpha_j^{(Av)} - \alpha_j^{(v)}$ for $j = 0, \ldots, s$. By construction, the η_j do not depend on γ. Moreover, we claim that $\eta_j = 0$ for $j = 0, \ldots, s$.

If $s = 0$, then the set of conditions (4.5) is empty and our claim follows directly from comparing (4.4) and (4.6). To show our claim for $1 \leq s \leq N - 2$, we rewrite (4.5) in the equivalent form

$$(4.7) \qquad \alpha_j^{(w)} = \alpha_j^{(v)} + \frac{1}{\gamma} (\alpha_{j-1}^{(Av)} - \alpha_{j-1}^{(w)}), \qquad j = 1, \ldots, s.$$

Then (4.6) and (4.7) for $j = s$ yield

$$\eta_s = \frac{1}{\gamma} (\alpha_{s-1}^{(Av)} - \alpha_{s-1}^{(w)}).$$

In this formula we can replace $\alpha_{s-1}^{(w)}$ by the right-hand side of (4.7) for $j = s - 1$,

$$\eta_s = \frac{1}{\gamma} \left(\alpha_{s-1}^{(Av)} - \alpha_{s-1}^{(v)} - \frac{1}{\gamma} (\alpha_{s-2}^{(Av)} - \alpha_{s-2}^{(w)}) \right) = \frac{1}{\gamma} \left(\eta_{s-1} - \frac{1}{\gamma} (\alpha_{s-2}^{(Av)} - \alpha_{s-2}^{(w)}) \right).$$

In the same way we now exploit (4.7) for $j = s - 2, \ldots, 1$, and finally use (4.4) to replace $\alpha_0^{(w)}$ by $\alpha_0^{(v)}$. The result of this process is equivalent to the relation

$$(4.8) \qquad \sum_{j=0}^{s} (-1)^{s-j} \eta_j \gamma^j = 0.$$

The coefficients η_j do not depend on γ, so that the left-hand side of (4.8) is a polynomial in γ of degree at most s. Since γ is allowed to vary almost freely without violating the assumption $d(A, v) = d(A, Av) = d(A, w) = N$ (see above), but on the other hand (4.8) must always hold, we conclude that $\eta_j = 0$ for $j = 0, \ldots, s$.

To summarize, since $d(A, v) = d(A, Av) = N$, (4.1) implies that $A^+ v = p_s(A)v$ and $A^+ Av = q_s(A)Av$ for two polynomials p_s and q_s of degree at most s, respectively. But since we have just shown that $p_s = q_s$, we receive

$$A^+ Av = q_s(A)Av = p_s(A)Av = Ap_s(A)v = AA^+ v.$$

Since we can find N linearly independent vectors v for which this is true, we conclude that $A^+ A = AA^+$, and indeed A must be B-normal(s). $\qquad \square$

Remark 4.2. This lemma represents a strengthened version of a result of Faber and Manteuffel [4, Lemma 2]. In their result, (4.1) is replaced by "$A^+ v_1 \in \mathcal{K}_{s+1}(A, v_1)$ for all v_1," and their proof of necessity uses an eigenvector v_1 of A, i.e., a vector v_1 with $d(A, v_1) = 1$. Our proof is inspired by an idea of Voevodin and Tyrtyshnikov [12].

Necessity in Theorem 3.2. We now come to the main goal of this section, namely the proof of necessity in Theorem 3.2. Our supposedly necessary condition reads *either A is B-normal(s), or $N - 1 \leq s$*. We will prove this by assuming the opposite and then showing that A is *not* B-reducible to $(s+2)$-band Hessenberg form.

The opposite of our necessary condition is that *A is not B-normal(s) and $0 \leq s \leq N - 2$*. But this is precisely the opposite of the nontrivial necessary condition in Lemma 4.1. Therefore, Lemma 4.1 implies that for each HPD matrix B there exists at least one vector v_1 with $d(A, v_1) = N$ such that $A^+ v_1 \notin \mathcal{K}_{s+1}(A, v_1)$. On the other hand, since $d(A, v_1) = N$, $A^+ v_1 \in \mathcal{K}_N(A, v_1)$. Hence $A^+ v_1$ is a linear combination of the basis vectors $v_1, \ldots, v_N$ computed by (2.5),

$$A^+ v_1 = \sum_{j=1}^{N} \beta_j v_j,$$

where (at least) one of the coefficients β_j, $s + 2 \leq j \leq N$, is nonzero. Let this be the coefficient with index k. Then, according to (2.6) and the B-orthogonality conditions (2.4),

$$(4.9) \qquad h_{1k} \; = \; \frac{\langle Av_k, v_1 \rangle_B}{\langle v_1, v_1 \rangle_B} \; = \; \frac{\langle v_k, A^+ v_1 \rangle_B}{\langle v_1, v_1 \rangle_B} \; = \; \frac{\overline{\beta}_k \langle v_k, v_k \rangle_B}{\langle v_1, v_1 \rangle_B} \; \neq \; 0 \,,$$

for an index k with $s + 2 \leq k \leq N$. Consequently, A is *not* B-reducible to $(s+2)$-band Hessenberg form, which completes the proof of necessity.

5. The existence of $(s + 2)$-term recurrences. We next relate Theorem 3.2 to the existence of an $(s + 2)$-term recurrence for computing B-orthogonal Krylov subspace bases.

Suppose that, for a given nonsingular matrix A, vector v_1, and HPD matrix B, only the previous $s + 1$ vectors $v_k, v_{k-1}, \ldots, v_{k-s}$ are required to compute v_{k+1}, $k = 1, \ldots, d - 1$, in (2.5). Then we say that the B-orthogonal basis $v_1, \ldots, v_d$ of $\mathcal{K}_d(A, v_1)$ is computed by an $(s + 2)$-term recurrence.

Note that the basic algorithm (2.5)–(2.6) has computed all basis vectors $v_1, \ldots, v_d$ in step $k = d - 1$. Therefore, in terms of the matrix representation (2.8), the B-orthogonal basis vectors are computed by an $(s + 2)$-term recurrence, if $H_{d,d-1}$ is $(s+2)$-band Hessenberg. We stress that, unlike for the B-reducibility to $(s+2)$-band Hessenberg form, we here use $H_{d,d-1}$ and not H_d. It is a subtle and easily overlooked fact that the last column of H_d plays no role for the computation of the basis vectors $v_1, \ldots, v_d$. This column indeed may be full, and still the basis vectors are computed by an $(s + 2)$-term recurrence.

DEFINITION 5.1. *The nonsingular matrix A admits an $(s+2)$-term recurrence, if there exists an HPD matrix B so that for each v_1, either $H_{d,d-1}$ in the decomposition (2.8) is $(s+2)$-band Hessenberg, or $d \leq s + 2$.*

The difference in the trivial conditions between this definition and Definition 2.1 ($d \leq s+2$ versus $d \leq s+1$) precisely corresponds to the different roles of the matrices $H_{d,d-1}$ and H_d. We immediately realize that if a nonsingular matrix A is B-reducible to $(s + 2)$-band Hessenberg form, then it also admits an $(s + 2)$-term recurrence. The reverse implication, however, does not hold. In other words, B-reducibility to $(s + 2)$-band Hessenberg form and admissibility of an $(s + 2)$-term recurrence are in general *not equivalent.* As an example, consider any nonsingular 3 by 3 matrix A with $d(A) = 3$. Then each v_1 with $d = 3$ leads to an H_3 of the form

$$H_3 \; = \; \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ 1 & h_{22} & h_{23} \\ 0 & 1 & h_{33} \end{bmatrix} .$$

Trivially A admits a 3-term recurrence as each $H_{3,2}$ has only 3 nonzero bands (here $s = 1$). But B-reducibility to 3-band Hessenberg form requires that $h_{13} = 0$ for all v_1. Since $d(A) = 3 > s + 1$, this holds by Theorem 3.2 only when A is B-normal(1).

Necessary and sufficient conditions for admissibility of an $(s + 2)$-term recurrence were proven in the fundamental paper of Faber and Manteuffel [4].

THEOREM 5.2 (Faber–Manteuffel theorem). *The nonsingular matrix A admits an $(s + 2)$-term recurrence if and only if either A is B-normal(s), or $d(A) \leq s + 2$.*

The proof of this result is based on a highly nontrivial and clever construction, which, unfortunately, provides little insight into the necessity of B-normality. A similar result was announced by Voevodin [11], but its proof by Voevodin and Tyrtyshnikov [12] is difficult to understand and appears to be unknown even to many

specialists in the field. Comparing Theorem 3.2 with the Faber–Manteuffel theorem yields the following important observation.

COROLLARY 5.3. *The nonsingular matrix A with $d(A) > s + 2$ is B-reducible to $(s+2)$-band Hessenberg form if and only if it admits an $(s+2)$-term recurrence.*

In other words, in all cases of practical interest; i.e., when $d(A)$ is "large" and s is "small," the two matrix properties studied in this paper in fact are equivalent.

6. Concluding remarks. The reader may now ask if we successfully tried to prove necessity in the Faber–Manteuffel theorem using similar elementary (linear algebra) means as for our Theorem 3.2. The answer is yes, we tried, but no, we were unsuccessful. To explain the main difficulty, at least in our opinion, consider our proof of necessity in section 4. We assume the opposite of the necessary conditions and construct a certain nonzero entry h_{1k}, $s + 2 \leq k \leq N$, in the first row of H_N, cf. (4.9). Hence H_N cannot be $(s + 2)$-band Hessenberg, which leads to a contradiction showing that the conditions are indeed necessary. Except for the range $s+2 \leq k \leq N$, we have no information about the location of the nonzero entry h_{1k}. If we could show that indeed $h_{1k} \neq 0$ for a k in the range $s + 2 \leq k \leq N - 1$, then $H_{N,N-1}$ cannot be $(s + 2)$-band Hessenberg either. This could subsequently be used to show necessity in the Faber–Manteuffel theorem. However, it is apparently quite difficult to "fix" k inside the range $s + 2 \leq k \leq N - 1$.

REFERENCES

[1] T. BARTH AND T. MANTEUFFEL, *Multiple recursion conjugate gradient algorithms Part* I: *Sufficient conditions*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 768–796.

[2] B. A. CIPRA, *The Best of the* 20*th Century: Editors Name Top* 10 *Algorithms*, SIAM News, 33 (2000).

[3] L. ELSNER AND K. D. IKRAMOV, *Normal matrices: An update*, Linear Algebra Appl., 285 (1998), pp. 291–303.

[4] V. FABER AND T. MANTEUFFEL, *Necessary and sufficient conditions for the existence of a conjugate gradient method*, SIAM J. Numer. Anal., 21 (1984), pp. 352–362.

[5] V. FABER AND T. A. MANTEUFFEL, *Orthogonal error methods*, SIAM J. Numer. Anal., 24 (1987), pp. 170–187.

[6] B. FISCHER, A. RAMAGE, D. J. SILVESTER, AND A. J. WATHEN, *Minimum residual methods for augmented systems*, BIT, 38 (1998), pp. 527–543.

[7] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[8] R. GRONE, C. R. JOHNSON, E. M. DE SÁ, AND H. WOLKOWICZ, *Normal matrices*, Linear Algebra Appl., 87 (1987), pp. 213–225.

[9] W. D. JOUBERT AND D. M. YOUNG, *Necessary and sufficient conditions for the simplification of generalized conjugate-gradient algorithms*, Linear Algebra Appl., 88/89 (1987), pp. 449–485.

[10] D. KHAVINSON AND G. ŚWIĄTEK, *On the number of zeros of certain harmonic polynomials*, Proc. Amer. Math. Soc., 131 (2003), pp. 409–414.

[11] V. V. VOEVODIN, *The question of non-self-adjoint extension of the conjugate gradients method is closed*, U.S.S.R. Comput. Math. and Math. Phys., 23 (1983), pp. 143–144.

[12] V. V. VOEVODIN AND E. E. TYRTYSHNIKOV, *On generalization of conjugate direction methods*, in Numerical Methods of Algebra (Chislennye Metody Algebry), Moscow State University Press, Moscow, 1981, pp. 3–9. English translation provided by E. E. Tyrtyshnikov.