The Faber–Manteuffel Theorem for Linear Operators

Article *in* SIAM Journal on Numerical Analysis · January 2008

DOI: 10.1137/060678087 · Source: DBLP

3 authors:

Vance Faber
Institute for Defense Analyses
127 PUBLICATIONS **3,962** CITATIONS
SEE PROFILE

Jörg Liesen
Technische Universität Berlin
120 PUBLICATIONS **2,225** CITATIONS
SEE PROFILE

Petr Tichý
Charles University in Prague
33 PUBLICATIONS **277** CITATIONS
SEE PROFILE

THE FABER-MANTEUFFEL THEOREM FOR LINEAR OPERATORS

V. FABER*, J. LIESEN†, AND P. TICHÝ‡

Abstract. A short recurrence for orthogonalizing Krylov subspace bases for a matrix A exists if and only if the adjoint of A is a low degree polynomial in A (i.e. A is normal of low degree). In the area of iterative methods, this result is known as the Faber-Manteuffel Theorem (V. Faber and T. Manteuffel, SIAM J. Numer. Anal., 21 (1984), pp. 352–362). Motivated by the description in (J. Liesen and Z. Strakoš, On optimal short recurrences for generating orthogonal Krylov subspace bases, SIAM Rev., to appear), we here formulate this theorem in terms of linear operators on finite dimensional Hilbert spaces, and give two new proofs of the necessity part. We have chosen the linear operator rather than the matrix formulation because we found that a matrix-free proof is less technical. Of course, the linear operator result contains the Faber-Manteuffel Theorem for matrices.

Key words. cyclic subspaces, Krylov subspaces, orthogonal bases, orthogonalization, short recurrences, normal matrices.

AMS subject classifications. 65F10, 65F25.

1. Introduction.

At the Householder Symposium VIII held in Oxford in July 1981, Golub posed as an open question to characterize necessary and sufficient conditions on a matrix A for the existence of a three-term conjugate gradient type method for solving linear systems with A (cf. SIGNUM Newsletter, vol. 16, no. 4, 1981). This important question was answered by Faber and Manteuffel in 1984 [4]. They showed that an $(s+2)$-term conjugate gradient type method for A, based on some given inner product exists if and only if the adjoint of A with respect to the inner product is a polynomial of degree s in A (i.e. A is normal of degree s). In the area of iterative methods this result is known as the Faber-Manteuffel Theorem; see, e.g., [7, Chapter 6] or [13, Chapter 6.10].

The theory of [4] and some further developments have recently been surveyed in [12]. There the Faber-Manteuffel Theorem is formulated independently of the conjugate gradient context, and solely as a result on the existence of a short recurrence for generating orthogonal bases for Krylov subspaces of the matrix A. A new proof of the sufficiency part is given, and the normality condition on A is thoroughly characterized. For the proof of the (significantly more difficult) necessity part, however, the authors refer to [4]. In particular, they suggest that, in light of the fundamental nature of the result, it is desirable to find an alternative, and possibly simpler proof.

Motivated by the description in [12], we here take a new approach to formulate and prove the necessity part of the Faber-Manteuffel Theorem. Instead of a matrix we consider a given linear operator A on a finite dimensional Hilbert space V. By the cyclic decomposition theorem, the space V decomposes into cyclic invariant subspaces, i.e. Krylov subspaces, of A (see Section 2 for details). The Faber-Manteuffel Theorem then gives a necessary (and sufficient) condition on A, so that the standard Gram-Schmidt algorithm for generating orthogonal bases of the cyclic subspaces reduces

*BD Biosciences-Bioimaging Systems, email: vance_faber@bd.com.

†Institute of Mathematics, Technical University of Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany, email: liesen@math.tu-berlin.de. The work of this author was supported by the Emmy Noether-Programm of the Deutsche Forschungsgemeinschaft. (Corresponding author)

‡Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 18207 Prague, Czech Republic, email: tichy@cs.cas.cz. The work of this author was supported by the Emmy Noether-Programm of the Deutsche Forschungsgemeinschaft and by the Czech National Program of Research "Information Society" under project 1ET400300415.

from a full to a short recurrence.

We have chosen this setting because we believe that the proof of necessity is easier to follow when we use linear operators rather than matrices. In this paper we give two different proofs of the necessity part, both based on restriction of the linear operator A to certain cyclic invariant subspaces. The resulting technicalities in the matrix formulation would obstruct rather than help the understanding. Moreover, our formulation may serve as a starting point for extending the results to infinite dimensional spaces. We are not aware that any such extensions have been obtained yet.

The paper is organized as follows. In Section 2 we introduce the notation and the required background from the theory of linear operators. In Section 3 we translate the matrix concepts introduced in [12] into the language of linear operators. In Section 4 we state and prove several technical lemmas that are required in the proof of the main result, which is given in Section 5. In Section 6 we give an alternative proof, which we consider elementary and constructive. This proof involves structure-preserving orthogonal transformations of Hessenberg matrices, which may be of interest beyond our context here. In Section 7 we discuss our rather theoretical analysis in the preceeding sections. This discussion includes a matrix formulation of the Faber-Manteuffel Theorem, a "high-level" description of the strategies of our two proofs of the necessity part, and our reasoning why necessity is more difficult to prove than sufficiency. For obtaining a more detailed overview of the results in this paper, Section 7 may also be read before reading the other sections.

2. Notation and background. In this section we introduce the notation and recall some basic results from the theory of linear operators; see Gantmacher's book [6, Chapters VII and IX] for more details.

Let V be a finite dimensional Hilbert space, i.e., a complex vector space equipped with a (fixed) inner product $(\cdot,\cdot)$. Let $A : V \to V$ be a given invertible linear operator. For any vector $v \in V$, we can form the sequence

$$v, Av, A^2v, \ldots . \tag{2.1}$$

Since V is finite dimensional, there exists an integer $d = d(A, v)$, such that the vectors $v, Av, \ldots, A^{d-1}v$ are linearly independent, while $A^d v$ is a linear combination of them. This means that there exist scalars, $\alpha_1, \ldots, \alpha_{d-1}$, not all equal to zero, such that

$$A^d v \;=\; -\sum_{j=0}^{d-1} \alpha_j A^j v\,. \tag{2.2}$$

Defining the monic polynomial $\phi(z) = z^d + \alpha_{d-1} z^{d-1} + \cdots + \alpha_0$, we can rewrite (2.2) as

$$\phi(A)v \;=\; 0\,. \tag{2.3}$$

We say that ϕ *annihilates* v. It would be more accurate to say "ϕ annihilates v with respect to A", but when it is clear which operator A is meant, the reference to A is omitted for the sake of brevity. The monic polynomial ϕ is the uniquely determined monic polynomial of smallest degree that annihilates v, and it is called the *minimal polynomial of* v. Its degree, equal to $d(A, v)$, is called the *grade of* v, and v is said to be of grade $d(A, v)$.

Consider any basis of V, and define the polynomial Φ as the least common multiple of the minimal polynomials of the basis vectors. Then Φ is the uniquely defined (independent of the choice of the basis!) monic polynomial of smallest degree that annihilates all vectors $v \in V$, and it is called the *minimal polynomial of* A. We denote its degree by $d_{\min}(A)$. Apparently, $d_{\min}(A) \geq d(A, v)$ for all $v \in V$, and Φ is divisible by the minimal polynomial of every vector $v \in V$.

If $v \in V$ is any vector of grade d, then

$$\operatorname{span}\{v, \dots A^{d-1}v\} \;\equiv\; \mathcal{K}_d(A, v) \tag{2.4}$$

is a d-dimensional invariant subspace of A. Because of (2.2) and the special character of the basis vectors, the subspace $\mathcal{K}_d(A, v)$ is called *cyclic*. The letter $\mathcal{K}$ has been chosen because this space is often called the *Krylov subspace* of A and v. The vector v is called the *generator* of this subspace.

A central result in the theory of linear operators on finite dimensional vector spaces is that the space V can be decomposed into cyclic subspaces. This result has several equivalent formulations, and in this paper we will use the one from [6, Chapter VII, §4, Theorem 3]: There exist vectors $w_1, \dots, w_j \in V$ of respective grades $d_1, \dots, d_j$, such that

$$V \;=\; \mathcal{K}_{d_1}(A, w_1) \oplus \cdots \oplus \mathcal{K}_{d_j}(A, w_j)\,, \tag{2.5}$$

where the minimal polynomial of w_1 is equal to the minimal polynomial of A, and for $k = 1, \dots, j-1$, the minimal polynomial of w_k is divisible by the minimal polynomial of w_{k+1}.

Since the decomposition (2.5) is an important tool in this paper, we illustrate it by a simple example (adapted from [9, Section 7.2]; also see [10] for a short and self-contained proof of the decomposition (2.5)). Suppose that A is the linear operator on $V = \mathbb{R}^3$ whose matrix representation in the canonical basis of $\mathbb{R}^3$ is

$$\begin{bmatrix} 2 & -3 & -3 \\ -3 & 2 & 3 \\ 3 & -3 & -4 \end{bmatrix}.$$

The characteristic polynomial of A is $(z-2)(z+1)^2$, while the minimal polynomial is $\Phi = (z-2)(z+1)$, so that $d_{\min}(A) = 2$. Any nonzero vector in $\mathbb{R}^3$ is either of grade one (and hence is an eigenvector) or of grade two. It is easy to see that the first canonical basis vector is not an eigenvector. Thus, $w_1 \equiv [1,\, 0,\, 0]^T$ is of grade $d_1 = 2$, i.e. $\mathcal{K}_{d_1}(A, w_1)$ has dimension two, and the minimal polynomial of w_1 is Φ. Note that

$$\mathcal{K}_{d_1}(A, w_1) \;=\; \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix} \right\} \;=\; \left\{ \begin{bmatrix} \alpha \\ \beta \\ -\beta \end{bmatrix} \,:\, \alpha, \beta \in \mathbb{R} \right\}.$$

Since $V = \mathbb{R}^3$ has dimension three, it remains to find a vector $w_2 \notin \mathcal{K}_{d_1}(A, w_1)$ that is of grade one and has minimal polynomial $z+1$, i.e. w_2 is an eigenvector with respect to the eigenvalue -1, that is not contained in $\mathcal{K}_{d_1}(A, w_1)$. These requirements are satisfied by $w_2 \equiv [1,\, 0,\, 1]^T$, giving

$$\mathbb{R}^3 \;=\; \mathcal{K}_{d_1}(A, w_1) \oplus \mathcal{K}_{d_2}(A, w_2) \;=\; \operatorname{span}\{w_1,\, Aw_1\} \oplus \operatorname{span}\{w_2\}\,.$$

In the basis of $\mathbb{R}^3$ given by w_1, Aw_1, w_2, the linear operator A has the matrix representation

$$\left[\begin{array}{cc|c} 0 & 2 & \\ 1 & 1 & \\ \hline & & -1 \end{array}\right].$$

Here the two diagonal blocks correspond to the decomposition (2.5), where each block is the companion matrix of the minimal polynomial of the respective cyclic subspace generators. This matrix representation is sometimes called the *rational canonical form*. When this canonical form consists of a single diagonal block in companion form, A is called *non-derogatory*. Hence in our example A is derogatory, but the restriction of A to the cyclic subspace generated by w_1 is non-derogatory. Loosely speaking, this restriction is the "largest non-derogatory part" of A.

3. Orthogonalization of a cyclic subspace basis. Let $v \in V$ be a vector of grade d. For theoretical as well as practical purposes it is often convenient to *orthogonalize* the basis $v, \ldots, A^{d-1}v$ of the cyclic subspace $\mathcal{K}_d(A, v)$. The classical approach to orthogonalization, which appears in different mathematical areas, cf., e.g., [2, p. 15], [5, p. 74], is the Gram-Schmidt algorithm:

$$v_1 = v\,, \tag{3.1}$$

$$v_{n+1} = Av_n - \sum_{m=1}^{n} h_{m,n} v_m\,, \tag{3.2}$$

$$h_{m,n} = \frac{(Av_n, v_m)}{(v_m, v_m)}\,, \quad m = 1, \ldots, n\,, \quad n = 1, \ldots, d-1\,. \tag{3.3}$$

The resulting vectors $v_1, \ldots, v_d$ are mutually orthogonal, and for $n = 1, \ldots, d$ they satisfy $\mathrm{span}\{v_1, \ldots, v_n\} = \mathrm{span}\{v, \ldots, A^{n-1}v\}$. We call v (or v_1) the *initial vector* of the algorithm (3.1)–(3.3). When A is a (square) matrix, this algorithm is usually referred to as Arnoldi's method [1]. It can be equivalently written as

$$v_1 = v\,, \tag{3.4}$$

$$A \underbrace{[v_1, \ldots, v_{d-1}]}_{\equiv V_{d-1}} = \underbrace{[v_1, \ldots, v_d]}_{\equiv V_d} \underbrace{\begin{bmatrix} h_{1,1} & \cdots & h_{1,d-1} \\ 1 & \ddots & \vdots \\ & \ddots & h_{d-1,d-1} \\ & & 1 \end{bmatrix}}_{\equiv H_{d,d-1}}, \tag{3.5}$$

$$(v_i, v_j) = 0 \quad \text{for } i \neq j\,, \quad i, j = 1, \ldots, d\,. \tag{3.6}$$

The matrix $H_{d,d-1}$ is an unreduced upper Hessenberg matrix of size $d \times (d-1)$. Its band structure determines the length of the recurrence (3.2) that generates the orthogonal basis. To state this formally, we need the following definition, cf. [12, Definition 2.1].

DEFINITION 3.1. *An unreduced upper Hessenberg matrix is called $(s+2)$-band Hessenberg, when its s-th superdiagonal contains at least one nonzero entry, and all its entries above its s-th superdiagonal are zero.*

If $H_{d,d-1}$ is $(s+2)$-band Hessenberg, then for $n = 1, \ldots, d-1$, the recurrence (3.2) reduces to

$$v_{n+1} \;=\; Av_n \;-\; \sum_{m=n-s}^{n} h_{m,n} v_m \,, \tag{3.7}$$

and thus the orthogonal basis is generated by an $(s+2)$-*term recurrence*. Since precisely the latest $s+1$ basis vectors $v_n, \ldots, v_{n-s}$ are required to determine v_{n+1}, and only one operation with A is performed, an $(s+2)$-term recurrence of the form (3.7) is called *optimal*.

Definition 3.2. *(Linear operator version of [12, Definition 2.4].) Let A be an invertible linear operator with minimal polynomial degree $d_{\min}(A)$ on a finite dimensional Hilbert space, and let s be a nonnegative integer, $s+2 \le d_{\min}(A)$.*

(1) If for an initial vector the matrix $H_{d,d-1}$ in (3.4)–(3.6) is $(s+2)$-band Hessenberg, then we say that A admits for the given initial vector an optimal $(s+2)$-term recurrence.

(2) If A admits for any given initial vector an optimal recurrence of length at most $s+2$, while it admits for at least one given initial vector an optimal $(s+2)$-term recurrence, then we say that A admits an optimal $(s+2)$-term recurrence.

We denote the *adjoint of* A by A^*. This is the uniquely determined operator that satisfies $(Av, w) = (v, A^*w)$ for all vectors v and w in the given Hilbert space. The operator A is called *normal* if it commutes with its adjoint, $AA^* = A^*A$. This holds if and only if A has a complete orthonormal system of eigenvectors. Equivalently, A^* can be written as a polynomial in A, $A^* = p(A)$, where p is completely determined by the condition that $p(\lambda_j) = \overline{\lambda}_j$ for all eigenvalues λ_j of A (cf. [6, Chapter IX, § 10]). We will be particularly interested in the degree of this polynomial.

Definition 3.3. *Let A be an invertible linear operator on a finite dimensional Hilbert space. If the adjoint of A satisfies $A^* = p(A)$, where p is a polynomial of smallest degree s having this property, then A is called normal of degree s, or, shortly, normal(s).*

The condition that A is normal(s) is *sufficient* for A to admit an optimal $(s+2)$-term recurrence. The precise formulation of this statement is the following.

Theorem 3.4. *Let A be an invertible linear operator with minimal polynomial degree $d_{\min}(A)$ on a finite dimensional Hilbert space. Let s be a nonnegative integer, $s+2 < d_{\min}(A)$. If A is normal(s), then A admits an optimal $(s+2)$-term recurrence.*

Proof. A matrix version of this result is given in [12, Theorem 2.9], and the proof given there can be easily adapted from matrices to linear operators. □

The main result we will prove in this paper is that the condition that A is normal(s) also is *necessary*.

Theorem 3.5. *Let A be an invertible linear operator with minimal polynomial degree $d_{\min}(A)$ on a finite dimensional Hilbert space. Let s be a nonnegative integer, $s+2 < d_{\min}(A)$. If A admits an optimal $(s+2)$-term recurrence, then A is normal(s).*

4. Technical lemmas. We prove Theorem 3.5 in Section 5 below. To do so, we need several technical lemmas that are stated and proven in this section.

LEMMA 4.1. *Let A be an invertible linear operator with minimal polynomial degree $d_{\min}(A)$ on a finite dimensional Hilbert space. If $1 < i < n \leq d_{\min}(A)$ and $(u_1, Au_i) = 0$ for every initial vector u_1 of grade n, then $(v_1, Av_i) = 0$ for every initial vector v_1 of grade m, where $i \leq m \leq n$.*
(Here u_i, v_i are the i-th basis vectors generated by (3.1)–(3.3) with initial vectors u_1, v_1, respectively.)

Proof. If $m = n$, there is nothing to prove. Hence, suppose that $m < n$, and let v_1 be a vector of grade m, and u_1 be a vector of grade n, such that the minimal polynomial of v_1 divides the minimal polynomial of u_1. Define

$$x_1 \;\equiv\; x_1(\gamma) \;\equiv\; v_1 + \gamma u_1\,, \tag{4.1}$$

where γ is a scalar parameter. It is clear that, except for finitely many choices of γ, the vector x_1 is of grade n.

Suppose that γ has been chosen so that x_1 is of grade n, and consider the corresponding i-th basis vector x_i, where $1 < i \leq m$. By construction, $x_i = p(A)x_1$, where p is a polynomial of (exact) degree $i-1$. The vector x_i is defined uniquely (up to scaling) by the conditions

$$(A^j x_1, x_i) \;=\; (A^j x_1, p(A)x_1) \;=\; 0\,, \quad j = 0, \ldots, i-2\,.$$

The hypothesis

$$(x_1, Ax_i) = (x_1, Ap(A)x_1) \;=\; 0$$

gives one additional condition. We thus have i conditions that translate into i homogeneous linear equations for the i coefficients of the polynomial p. The existence of x_i implies that the determinant of the matrix $M(x_1)$ of these equations must be zero, where

$$M(x_1) \;=\; \begin{bmatrix} (x_1, x_1) & (x_1, Ax_1) & \cdots & (x_1, A^{i-1}x_1) \\ (Ax_1, x_1) & (Ax_1, Ax_1) & \cdots & (Ax_1, A^{i-1}x_1) \\ \vdots & \vdots & \vdots & \vdots \\ (A^{i-2}x_1, x_1) & (A^{i-2}x_1, Ax_1) & \cdots & (A^{i-2}x_1, A^{i-1}x_1) \\ (x_1, Ax_1) & (x_1, A^2x_1) & \cdots & (x_1, A^i x_1) \end{bmatrix}.$$

Now note that $\det M(x_1)$ is a continuous function of γ. By construction, this function is zero for all but a finite number of choices of γ. Therefore $\det M(x_1) = 0$ for all γ, and in particular, $\det M(v_1) = 0$. Consequently, there exists a non-trivial solution of the linear system with $M(v_1)$, defining a vector $w = p(A)v_1$, where p is a polynomial of degree at most $i-1$. The first $i-1$ rows mean that w is orthogonal to $v_1, \ldots, A^{i-2}v_1$, so w is a multiple of v_i. The last row means that Aw and hence Av_i is orthogonal to v_1. □

The decomposition (2.5) shows that for any linear operator A on a finite dimensional Hilbert space V, there exists a vector in V whose minimal polynomial coincides with the minimal polynomial of A. The following result shows that there in fact exists a basis of V consisting of vectors with this property.

LEMMA 4.2. *Let A be an invertible linear operator with minimal polynomial degree $d_{\min}(A)$ on a finite dimensional Hilbert space V. Then there exists a basis of V consisting of vectors that all are of grade $d_{\min}(A)$.*

Proof. From the cyclic decomposition theorem, cf. (2.5), we know that there exist vectors $w_1, \dots, w_j$ with minimal polynomials $\phi_1, \dots, \phi_j$ of respective degrees $d_1, \dots, d_j$, such that the space V can be decomposed as

$$V \;=\; \mathcal{K}_{d_1}(A, w_1) \oplus \cdots \oplus \mathcal{K}_{d_j}(A, w_j)\,,$$

where ϕ_1 equals the minimal polynomial of A, and ϕ_k is divisible by ϕ_{k+1} for $k = 1, \dots, j-1$. In particular, $d_1 = d_{\min}(A)$. Consequently, a basis of V is given by

$$w_1, \dots, A^{d_1-1}w_1, \quad w_2, \dots, A^{d_2-1}w_2, \quad \dots, \quad w_j, \dots, A^{d_j-1}w_j\,.$$

But then it is easy to see that

$$w_1, \dots, A^{d_1-1}w_1, \quad w_2 + w_1, \dots, A^{d_2-1}w_2 + w_1, \quad \dots, \quad w_j + w_1, \dots, A^{d_j-1}w_j + w_1$$

is a basis of V consisting of vectors that all are of grade d_1. ∎

The following result is a generalization of [11, Lemma 4.1], which in turn can be considered a (considerably) strengthened version of [4, Lemma 2].

LEMMA 4.3. *Let A be an invertible linear operator with minimal polynomial degree $d_{\min}(A)$ on a finite dimensional Hilbert space. Let B be a linear operator on the same space, and let s be a nonnegative integer, $s + 2 \le d_{\min}(A)$. If*

$$Bv \;\in\; \operatorname{span}\{v, \dots, A^s v\} \quad \textit{for all vectors } v \textit{ of grade } d_{\min}(A)\,, \tag{4.2}$$

then $AB = BA$. In particular, if $B = A^*$, then A is normal(t) for some $t \le s$.

Proof. For notational convenience, we denote $\delta = d_{\min}(A)$. Let v be any vector of grade δ. Since A is invertible, $\mathcal{K}_\delta(A, v) = \mathcal{K}_\delta(A, Av)$. In addition, except possibly when γ is an eigenvalue of A, the vector $w = (A - \gamma I)v$ satisfies $\mathcal{K}_\delta(A, w) = \mathcal{K}_\delta(A, v)$. In the following, we exclude those values of γ. By assumption, there exist polynomials p_γ, q, and r of degree at most s, which satisfy

$$Bw = p_\gamma(A)w, \quad B(Av) = q(A)(Av), \quad Bv = r(A)v,$$

where p_γ depends on γ, but q and r do not. We can then write Bw as

$$Bw \;=\; p_\gamma(A)w \;=\; p_\gamma(A)Av - \gamma p_\gamma(A)v\,,$$

and

$$Bw \;=\; BAv - \gamma Bv \;=\; q(A)Av - \gamma r(A)v\,.$$

Combining these two identities yields

$$t_\gamma(A)v \;=\; 0\,, \quad \text{where} \quad t_\gamma(z) = z(p_\gamma(z) - q(z)) - \gamma(p_\gamma(z) - r(z))\,.$$

The polynomial t_γ is of degree at most $s+1 < s+2 \le \delta$. Thus, except for finitely many γ, $t_\gamma = 0$. Some straightforward algebraic manipulation gives, for all but these γ,

$$\gamma(q(z) - r(z)) \;=\; (z - \gamma)\widehat{p}_\gamma(z)\,,$$

where $\widehat{p}_\gamma \equiv p_\gamma - q$ is of degree at most $s-1$. Therefore, every γ that is not an eigenvalue of A is a root of the polynomial $r-q$, which consequently must be identically zero.

But then

$$B(Av) \;=\; q(A)(Av) \;=\; Aq(A)v \;=\; Ar(A)v \;=\; A(Bv)\,.$$

By Lemma 4.2, there exists a basis consisting of vectors of grade δ. Hence $BAv = ABv$ for a basis of vectors v, so that $BA = AB$.

Finally, if $B = A^*$, then $AA^* = A^*A$, so that A is normal and hence $A^* = p(A)$ for some polynomial. From (4.2) we see that the degree of p is at most s, so that A is normal(t) for some $t \le s$. □

5. Proof of Theorem 3.5. Let A be an invertible linear operator on a finite dimensional Hilbert space, and let s be a nonnegative integer, $s+2 < d_{\min}(A)$. Suppose that A admits an optimal $(s+2)$-term recurrence.

Step 1: Restriction to a cyclic subspace of dimension $s+2$.
If u_1 is any vector of grade $s+3$, then (with the obvious meaning of u_{s+2})

$$0 \;=\; h_{1,s+2} \;=\; (u_1, Au_{s+2})\,. \tag{5.1}$$

Consider any v_1 of grade $s+2$, and the corresponding cyclic subspace $\mathcal{K}_{s+2}(A, v_1)$. Let $\widehat{A}$ be the restriction of A to $\mathcal{K}_{s+2}(A, v_1)$, i.e. the invertible linear operator

$$\widehat{A} : \mathcal{K}_{s+2}(A, v_1) \to \mathcal{K}_{s+2}(A, v_1)\,, \qquad v \mapsto Av \;\text{ for } v \in \mathcal{K}_{s+2}(A, v_1)\,.$$

Clearly, $d_{\min}(\widehat{A}) = s+2$. Let $\mathcal{K}_{s+2}(A, v_1)$ be equipped with the same inner product as the whole space.

Let $y_1 \in \mathcal{K}_{s+2}(A, v_1)$ be any vector of grade $s+2$. Obviously, the grade of y_1 with respect to A is the same as the grade of y_1 with respect to $\widehat{A}$. Since (5.1) holds for any u_1 of grade $s+3$ (with respect to A), Lemma 4.1 (with $i = m = s+2$ and $n = s+3$) implies that (with the obvious meaning of y_{s+2})

$$0 \;=\; (y_1, Ay_{s+2}) \;=\; (y_1, \widehat{A}y_{s+2}) \;=\; (\widehat{A}^* y_1, y_{s+2})\,, \tag{5.2}$$

where $\widehat{A}^* : \mathcal{K}_{s+2}(A, v_1) \to \mathcal{K}_{s+2}(A, v_1)$ is the adjoint operator of $\widehat{A}$. But this means that

$$\widehat{A}^* y_1 \;\in\; \operatorname{span}\{y_1, \ldots, \widehat{A}^s y_1\}\,. \tag{5.3}$$

Since this holds for any vector $y_1 \in \mathcal{K}_{s+2}(A, v_1) = \mathcal{K}_{s+2}(\widehat{A}, v_1)$ of grade $s+2 = d_{\min}(\widehat{A})$, Lemma 4.3 implies that $\widehat{A}$ is normal(t) for some $t \le s$. In particular, $\widehat{A}$ is normal, and has $s+2$ distinct eigenvalues, λ_k, $k = 1, \ldots, s+2$, with corresponding eigenvectors that are mutually orthogonal. Moreover, there exists a polynomial of degree at most s such that $p(\lambda_k) = \overline{\lambda}_k$, $k = 1, \ldots, s+2$. By definition, any eigenpair of $\widehat{A}$ is an eigenpair of A. Therefore, A acting on *any* vector of grade $s+2$ has $s+2$ distinct eigenvalues, and the corresponding eigenvectors are mutually orthogonal in the given inner product.

Step 2: Extension to the whole space.
Consider the cyclic decomposition of the whole space as in (2.5). Then the cyclic

subspace $\mathcal{K}_{d_1}(A, w_1)$, where w_1 has the same minimal polynomial as A, can be further decomposed into

$$\mathcal{K}_{d_1}(A, w_1) \;=\; \mathcal{K}_{c_1}(A, z_1) \oplus \cdots \oplus \mathcal{K}_{c_\ell}(A, z_\ell)\,,$$

where the minimal polynomial of z_k is $(z - \lambda_k)^{c_k}$, $k = 1, \ldots, \ell$, and $\lambda_1, \ldots, \lambda_\ell$ are the distinct eigenvalues of A (cf., e.g., [6, Chapter VII, §2, Theorem 1]). In other words, $\mathcal{K}_{d_1}(A, w_1)$ is decomposed into ℓ cyclic invariant subspaces of A, where each of these corresponds to one of the ℓ distinct eigenvalues of A (recall that the restriction of A to $\mathcal{K}_{d_1}(A, w_1)$ is non-derogatory; cf. the example at the end of Section 2). In particular, if A is diagonalizable, then $\ell = d_{\min}(A)$, and $c_1 = \cdots = c_\ell = 1$, and $z_1, \ldots, z_\ell$ are eigenvectors of A corresponding to $\lambda_1, \ldots, \lambda_\ell$, respectively. In general, we can assume that $c_1 \geq c_2 \geq \cdots \geq c_\ell$. If $c_1 \geq s + 2$, we can determine a vector v_1 of grade $s + 2$ in $\mathcal{K}_{c_1}(A, z_1)$. But then the above implies that A acting on v_1 has $s + 2$ distinct eigenvalues, which is a contradiction. Hence $c_1 < s + 2$. We therefore can find an index m so that $c_1 + \cdots + c_{m-1} + \tilde{c}_m = s + 2$, $0 \leq \tilde{c}_m \leq c_m$. Let $\tilde{z}_m$ be any vector of grade $\tilde{c}_m$ in $\mathcal{K}_{c_m}(A, z_m)$, then $w = z_1 + \cdots + z_{m-1} + \tilde{z}_m$ is of grade $s + 2$. Hence A acting on w has $s + 2$ distinct eigenvalues, which shows that $c_1 = c_2 = \cdots = c_\ell = 1$. To these eigenvalues correspond $s + 2$ eigenvectors that are mutually orthogonal in the given inner product.

In the cyclic decomposition (2.5), the minimal polynomial of w_k is divisible by the minimal polynomial of w_{k+1}. Therefore the whole space completely decomposes into one-dimensional cyclic subspaces of A, i.e. A has a complete system of eigenvectors. We know that any $s + 2$ of these corresponding to distinct eigenvalues of A must be mutually orthogonal. In the subspaces corresponding to a multiple eigenvalue we can find an orthogonal basis. Therefore A has a complete orthonormal system of eigenvectors, and hence A is normal. For every subset of $s+2$ distinct eigenvalues there exists a polynomial p of degree at most s that satisfies $p(\lambda_k) = \overline{\lambda}_k$ for all eigenvalues λ_k in the subset. If we take any two subsets having $s + 1$ eigenvalues in common, the two corresponding polynomials must be identical. Thus all the polynomials are identical, so that A is normal(t) for some $t \leq s$.

If $t < s$, then by the sufficiency result in Theorem 3.4, A admits an optimal $(t + 2)$-term recurrence, which contradicts our initial assumption. Hence $t = s$, so that A is normal(s), which concludes the proof.

6. Another proof based on the "Rotation Lemma". In this section we discuss an elementary and more constructive approach to proving Theorem 3.5, which is based on orthogonal transformations ("rotations") of upper Hessenberg matrices. With this approach, we can prove Theorem 3.5 with the assumption "$s+2 < d_{\min}(A)$" replaced by "$s + 3 < d_{\min}(A)$". We discuss the "missing case" $s + 3 = d_{\min}(A)$ in Section 7.

As above, let A be an invertible linear operator with minimal polynomial degree $d_{\min}(A)$ on a finite dimensional Hilbert space. Let s be a given nonnegative integer, $s + 3 < d_{\min}(A)$. We assume that

(6.1) A admits an $(s + 2)$-term recurrence, but A is *not* normal(s),

and derive a contradiction.

For deriving the contradiction we need some notation. Suppose that the space is decomposed into cyclic invariant subspaces of A as in (2.5). Let $\widehat{A}$ be the restriction

of A to $\mathcal{K}_{d_1}(A, w_1)$, i.e. the invertible linear operator defined by

$$\widehat{A} \,:\, \mathcal{K}_{d_1}(A, w_1) \to \mathcal{K}_{d_1}(A, w_1)\,, \qquad v \mapsto Av \;\text{ for } v \in \mathcal{K}_{d_1}(A, w_1)\,.$$

The operator $\widehat{A}$ depends on the choice of w_1, which we consider fixed here, so $\widehat{A}$ is fixed as well. It is clear that $d_1 = d_{\min}(A) = d_{\min}(\widehat{A})$. We denote $d = d_1$ for simplicity. Now let $v_1 \in \mathcal{K}_d(\widehat{A}, w_1)$ be any initial vector of grade d, and let $v_1, \ldots, v_d$ be the corresponding orthogonal basis of $\mathcal{K}_d(\widehat{A}, v_1) = \mathcal{K}_d(\widehat{A}, w_1)$ generated by (3.1)–(3.3). Then the matrix representation of the operator $\widehat{A}$ with respect to this particular basis is a $d \times d$ unreduced upper Hessenberg matrix H_d, which is defined by the equation

$$\widehat{A}\,[v_1, \ldots, v_d] \;=\; [v_1, \ldots, v_d]\, H_d\,. \tag{6.2}$$

The matrix formed by the first $d-1$ columns of H_d coincides with the $d \times (d-1)$ upper Hessenberg matrix generated by (3.1)–(3.3) with $\widehat{A}$ and the initial vector v_1, while the last column of H_d is given by the vector

$$h_d \;=\; \begin{bmatrix} h_{1,d} \\ \vdots \\ h_{d,d} \end{bmatrix}, \quad \text{where} \quad h_{m,d} \;=\; \frac{(\widehat{A}v_d, v_m)}{(v_m, v_m)}, \quad m = 1, \ldots, d\,. \tag{6.3}$$

In short, $H_d = [H_{d,d-1}, h_d]$. We now proceed in two steps.

Step 1: Show that there exists a basis for which $h_{1,d} \neq 0$.
We first show that under assumption (6.1) there exists an initial vector $v_1 \in \mathcal{K}_d(\widehat{A}, w_1)$ of grade $d = d_{\min}(\widehat{A})$ for which the matrix representation H_d of $\widehat{A}$ has $h_{1,d} \neq 0$. Suppose not, i.e., for all $v_1 \in \mathcal{K}_d(\widehat{A}, w_1)$ of grade $d_{\min}(\widehat{A})$, we have for the resulting entry $h_{1,d}$,

$$0 \;=\; h_{1,d} \;=\; \frac{(\widehat{A}v_d, v_1)}{(v_1, v_1)} \;=\; \frac{(v_d, \widehat{A}^* v_1)}{(v_1, v_1)}\,,$$

where $\widehat{A}^*$ is the adjoint of $\widehat{A}$. In particular, this implies that for all vectors $v_1 \in \mathcal{K}_d(\widehat{A}, w_1)$ of grade $d = d_{\min}(\widehat{A})$,

$$\widehat{A}^* v_1 \;\in\; \{v_1, \ldots, \widehat{A}^{d-2} v_1\}\,.$$

By Lemma 4.3, $\widehat{A}$ is normal(t) for some $t \leq d_{\min}(\widehat{A}) - 2$. Therefore, A acting on any vector of grade $d_{\min}(A)$ has $d_{\min}(A)$ distinct eigenvalues and corresponding eigenvectors that are mutually orthogonal. From this it is easy to see that A is normal(t). By the sufficiency result in Theorem 3.4, A admits an optimal $(t+2)$-term recurrence. However, we have assumed in (6.1) that A admits an optimal $(s+2)$-term recurrence, so $t = s$. But then A is normal(s), which contradicts the second part of the assumption. In summary, there exists an initial vector v_1 of grade $d = d_{\min}(A)$, such that (6.2) holds with $H_d = [H_{d,d-1}, h_d]$, where $H_{d,d-1}$ is $(s+2)$-band Hessenberg (this follows from the first part of our assumption), while $h_{1,d} \neq 0$.

Step 2. "Rotation" of the nonzero entry $h_{1,d}$.
The following result is called "Rotation Lemma" for reasons apparent from its proof.

LEMMA 6.1. *(Rotation Lemma) Let s, d be nonnegative integers, $s + 3 < d$. Let H_d be a $d \times d$ unreduced upper Hessenberg matrix with $h_{1,d} \neq 0$ and $H_{d,d-1}$, the matrix*

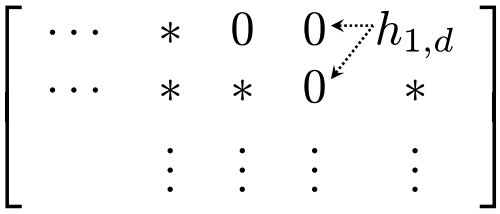

FIG. 6.1. *Graphical illustration of the "Rotation Lemma": Shown is the upper right hand corner of* $H_d = [H_{d,d-1}, h_d]$. *We know that* $H_{d,d-1}$ *is* $(s+2)$*-band Hessenberg with* $s+3 < d$, *and that* $h_{1,d} \neq 0$. *We construct an orthogonal transformation* G *such that the matrix* $\widetilde{H}_d = G^* H_d G$ *remains unreduced upper Hessenberg, while the nonzero entry* $h_{1,d} \neq 0$ *is "rotated" to the last column of* $\widetilde{H}_{d,d-1}$, *so that at least one of its entries* $\widetilde{h}_{1,d-1}$ *and* $\widetilde{h}_{2,d-1}$ *is nonzero.*

formed by the first $d-1$ *columns of* H_d, *being an* $(s+2)$*-band Hessenberg matrix. Then there exists a unitary matrix* G *such that* $\widetilde{H}_d \equiv G^* H_d G$ *is a* $d \times d$ *unreduced upper Hessenberg matrix with* $[\widetilde{h}_{1,d-1}, \widetilde{h}_{2,d-1}] \neq [0,0]$.

Proof. The main idea of this proof is to find $d-1$ (complex) Givens rotations of the form

$$G_i \equiv \begin{bmatrix} I_{d-1-i} & & & \\ & c_i & \overline{s}_i & \\ & -s_i & c_i & \\ & & & I_{i-1} \end{bmatrix}, \quad c_i^2 + |s_i|^2 = 1, \quad c_i \in \mathbb{R}, \quad i = 1, \ldots, d-1, \tag{6.4}$$

which, applied symmetrically to H_d, "rotate" the nonzero entry $h_{1,d}$ to the $(d-1)$st column of the resulting matrix $\widetilde{H}_d = (G_1 \cdots G_{d-1})^* H_d (G_1 \cdots G_{d-1})$. To prove the assertion it suffices to show the following: First, $\widetilde{H}_d$ must be an unreduced upper Hessenberg matrix, and, second, at least one of its entries $\widetilde{h}_{1,d-1}, \widetilde{h}_{2,d-1}$ is nonzero (see Fig. 6.1 for an illustration of this idea).

Proceeding in an inductive manner, we denote $H^{(0)} \equiv H_d$. To start, choose $s_1 \in \mathbb{R} \setminus \{0\}$ and $c_1 \in \mathbb{R}$ such that $c_1^2 + s_1^2 = 1$. We have explicitly chosen real parameters s_1, c_1 since this simplifies our arguments below. These two parameters determine our first Givens rotation G_1 of the form (6.4). By construction, the matrix $H^{(1)} \equiv G_1^* H^{(0)} G_1$ is upper Hessenberg except for its entry

$$h_{d,d-2}^{(1)} = s_1 h_{d-1,d-2}^{(0)}.$$

Since $s_1 \neq 0$ and $h_{d-1,d-2}^{(0)} \neq 0$ ($H^{(0)}$ is unreduced), we have $h_{d,d-2}^{(1)} \neq 0$. The transformation by G_1 modifies only the last two rows and columns of $H^{(0)}$, so that the entries on the subdiagonal of $H^{(1)}$ satisfy $h_{i+1,i}^{(1)} = h_{i+1,i}^{(0)} \neq 0$, $i = 1, \ldots, d-3$. Next, we determine G_2 such that its application from the right to $H^{(1)}$ eliminates the nonzero entry in position $(d, d-2)$. Application of G_2^* from the left then introduces a nonzero entry in position $(d-1, d-3)$, which we will subsequently eliminate using G_3, and so forth.

In a general step $j = 2, \ldots, d-1$, suppose that $s_{j-1} \neq 0$, $h_{i+1,i}^{(j-1)} = h_{i+1,i}^{(0)} \neq 0$, $i = 1, \ldots, d-j-1$, and $h_{i+1,i}^{(j-1)} \neq 0$ for $i = d-j+2, \ldots, d-1$. Next suppose that

$$H^{(j-1)} \equiv G_{j-1}^* H^{(j-2)} G_{j-1}$$

is an upper Hessenberg matrix except for its entry

$$h^{(j-1)}_{d-j+2,d-j} = s_{j-1} h^{(0)}_{d-j+1,d-j} \neq 0.$$

The next Givens rotation G_j is (uniquely) determined to eliminate this nonzero entry, i.e. we determine c_j and s_j by the equation

$$[h^{(j-1)}_{d-j+2,d-j}, h^{(j-1)}_{d-j+2,d-j+1}] \begin{bmatrix} c_j & \overline{s}_j \\ -s_j & c_j \end{bmatrix} = [0, h^{(j)}_{d-j+2,d-j+1}]. \tag{6.5}$$

Since $h^{(j-1)}_{d-j+2,d-j} \neq 0$, it is clear that $s_j \neq 0$ and $h^{(j)}_{d-j+2,d-j+1} \neq 0$. As a result, the matrix

$$H^{(j)} \equiv G_j^* H^{(j-1)} G_j$$

is an upper Hessenberg except for its entry

$$h^{(j)}_{d-j+1,d-j-1} = s_j h^{(0)}_{d-j,d-j-1} \neq 0.$$

The unitary transformation determined by G_j modifies only $(d-j)$th and $(d-j+1)$st rows and columns of $H^{(j-1)}$. Therefore, the subdiagonal entries of $H^{(j)}$ satisfy $h^{(j)}_{i+1,i} = h^{(0)}_{i+1,i} \neq 0$, for $i = 1, \ldots, d-j-2$, and, since $h^{(j)}_{d-j+2,d-j+1} \neq 0$, we have shown inductively that indeed $h^{(j)}_{i+1,i} \neq 0$ for $i = d-j+1, \ldots, d-1$. In the end, we receive the unitary matrix $G = G_1 \cdots G_{d-1}$ and the upper Hessenberg matrix $H^{(d-1)} = G^* H^{(0)} G$ with $h^{(d-1)}_{i+1,i} \neq 0$ for $i = 2, \ldots, d-1$. To complete the proof we need to show that the initial parameters s_1, c_1 can be chosen so that, first, $h^{(d-1)}_{2,1} \neq 0$ ($H^{(d-1)}$ is unreduced), and, second, $[\, h^{(d-1)}_{1,d-1}, h^{(d-1)}_{2,d-1}\,] \neq [0,0]$.

First, if $h^{(d-1)}_{2,1} = 0$, then we must have $h^{(d-1)}_{1,1} \neq 0$, for if otherwise $H^{(d-1)}$ would be singular. From $H^{(d-1)} = G^* H^{(0)} G$ we receive $H^{(0)} G = G H^{(d-1)}$, and thus the first column of G is an eigenvector of $H^{(0)}$ corresponding to the eigenvalue $h^{(d-1)}_{1,1}$. Note that the first column of G depends on our choice of s_1, while the matrix $H^{(0)}$ is fixed and has at most d linearly independent eigenvectors. Apparently, the case $h^{(d-1)}_{2,1} = 0$ only happens for a finite number of values of s_1 (if any); almost every initial choice of s_1 will yield $h^{(d-1)}_{2,1} \neq 0$.

Second, we have assumed that the first $d-1$ columns of $H^{(0)}$ form an unreduced $(s+2)$-band Hessenberg matrix with $s+3 < d$, and therefore $h^{(0)}_{1,d-2} = h^{(0)}_{1,d-1} = 0$ (cf. Fig. 6.1). Denote the entries of the (lower Hessenberg) matrix G by $g_{i,j}$. It is easy to see that $g_{d,d-1} = -c_2 s_1$. Again consider the matrix equation $H^{(0)} G = G H^{(d-1)}$. Comparing the entries in position $(1, d-1)$ on both sides shows that

$$-c_2 s_1 h^{(0)}_{1,d} \;=\; g_{1,1} h^{(d-1)}_{1,d-1} + g_{1,2} h^{(d-1)}_{1,d-1}\,, \tag{6.6}$$

where $h^{(0)}_{1,d} \neq 0$ and $s_1 \neq 0$. Therefore, to show that $[h^{(d-1)}_{1,d-1}, h^{(d-1)}_{2,d-1}] \neq [0,0]$, it suffices to show that $c_2 \neq 0$. For c_2 it holds that, cf. (6.5),

$$h^{(1)}_{d,d-2} c_2 - h^{(1)}_{d,d-1} s_2 \;=\; 0.$$

We know that $h^{(1)}_{d,d-2} \neq 0 \neq s_2$. Thus, $c_2 = 0$ if and only if $h^{(1)}_{d,d-1} = 0$, which holds if and only if

$$c_1 s_1 h^{(0)}_{d-1,d-1} + c_1^2 h^{(0)}_{d,d-1} - s_1^2 h^{(0)}_{d-1,d} - c_1 s_1 h^{(0)}_{d,d} \;=\; 0.$$

We write $s_1 = \sin(\theta)$, $c_1 = \cos(\theta)$, and apply standard identities for trigonometric functions to see that the above equation is equivalent with

$$\left(h^{(0)}_{d-1,d-1} - h^{(0)}_{d,d}\right)\sin(2\theta) + \left(h^{(0)}_{d,d-1} + h^{(0)}_{d-1,d}\right)\cos(2\theta) + \left(h^{(0)}_{d,d-1} - h^{(0)}_{d-1,d}\right) \;=\; 0\,.$$

The left hand side in this equation is a nontrivial trigonometric polynomial of degree two, which has at most two roots in the interval $[0, 2\pi)$. Consequently, for almost all choices of s_1 we receive $c_2 \neq 0$, giving a nonzero right hand side in (6.6). Hence, for almost all choices of s_1, we must have $[h^{(d-1)}_{1,d-1}, h^{(d-1)}_{2,d-1}] \neq [0, 0]$. □

We can now derive the contradiction to (6.1). Consider the relation (6.2), where H_d is of the form assumed in the Lemma 6.1. Without loss of generality we may assume that the columns of V_d are normalized (normalization does not alter the nonzero pattern of H_d). By Lemma 6.1, there exists a unitary matrix G such that $\widetilde{H}_d = G^* H_d G$ is unreduced upper Hessenberg with either $\widetilde{h}_{1,d-1}$ or $\widetilde{h}_{2,d-1}$ nonzero. Then (6.2) is equivalent with

$$\widehat{A}(V_d G) \;=\; (V_d G)\widetilde{H}_d\,. \tag{6.7}$$

Denote the entries of G by $g_{i,j}$, and let $V_d G \equiv [y_1, \ldots, y_d]$. Then, since the basis $v_1, \ldots, v_d$ is orthonormal and the matrix G is unitary, the basis $y_1, \ldots, y_d$ is orthonormal,

$$(y_i, y_j) \;=\; \left(\sum_{k=1}^{d} v_k g_{k,i}, \sum_{k=1}^{d} v_k g_{k,j}\right) \;=\; \sum_{k=1}^{d} \overline{g}_{k,j} g_{k,i} \;=\; \delta_{i,j}\,,$$

where $\delta_{i,j}$ is the Kronecker delta. By (6.7), the vectors $y_1, \ldots, y_d$ form the unique (up to scaling) basis of $\mathcal{K}_d(\widehat{A}, y_1)$ generated by (3.1)–(3.3) with $\widehat{A}$ and starting vector y_1. But since $[\widetilde{h}_{1,d-1}, \widetilde{h}_{2,d-1}] \neq [0, 0]$, we see that $\widehat{A}$ (and hence A) admits for the given y_1 an optimal recurrence of length at least $d - 1$. Since we have assumed that A admits an optimal $(s + 2)$-term recurrence, we must have $d - 1 \leq s + 2$, or, equivalently, $d = d_{\min}(A) \leq s + 3$. This is a contradiction since $s + 3 < d_{\min}(A)$.

As claimed at the beginning of this section, we now have shown Theorem 3.5, with the assumption "$s + 2 < d_{\min}(A)$" replaced by "$s + 3 < d_{\min}(A)$".

7. Concluding discussion. In this section we discuss our rather theoretical analysis above.

1. Matrix formulation and the Faber-Manteuffel Theorem.
When formulated in terms of matrices rather than linear operators, Theorems 3.4 and 3.5 comprise the Faber-Manteuffel Theorem [4] in the formulation given in [12, Section 2]. We state this result here for completeness.

THEOREM 7.1. *Let A be an $N \times N$ nonsingular matrix with minimal polynomial degree $d_{\min}(A)$. Let B be an $N \times N$ Hermitian positive definite matrix, and let s be a nonnegative integer, $s + 2 < d_{\min}(A)$. Then A admits for the given B an optimal $(s + 2)$-term recurrence if and only if A is B-normal(s).*

In this formulation, the Hilbert space from Theorems 3.4 and 3.5 is $\mathbb{C}^N$, equipped with the inner product generated by the Hermitian positive definite matrix B (in case A is real, we consider B to be real as well, and the adjoint A^* is the regular transpose A^T). The matrix A is B-normal(s) if its B-adjoint, i.e. the matrix $A^+ \equiv B^{-1}A^*B$,

is a polynomial of degree s in A, and s is the smallest degree for which this is true. A complete characterization of the matrices A and B for which A is B-normal(s) is given in [12, Section 3].

In this paper we have chosen the linear operator rather than the matrix formulation, because it appears to be a natural generalization. Moreover, both proofs we have given use the restriction of the linear operator A to certain cyclic invariant subspaces. In the matrix formulation, such restrictions lead to non-square as well as square but singular matrices. This involves a more complicated notation, which obstructs rather than helps the theoretical understanding. For instance, the restriction $\widehat{A}$ of a nonsingular $N \times N$ matrix A to a cyclic invariant subspace of A with (orthonormal) basis $v_1, \ldots, v_d$ can be represented as $\widehat{A} = VHV^*$, where $V = [v_1, \ldots, v_d]$ and H is a $d \times d$ nonsingular matrix. If $d < N$, $\widehat{A}$ is a singular $N \times N$ matrix (more precisely, $\widehat{A}$ has rank $d < N$). Any vector w in the cyclic invariant subspace can be represented as $w = V\omega$, where ω is a vector of length d containing the coefficients of w in the basis, so that $Aw = \widehat{A}w = VH\omega$, where VH is a (non-square) matrix of size $N \times d$. On the other hand, in the linear operator formulation, $\widehat{A}$ is invertible, and we may simply write $\widehat{A}w$ for the application of $\widehat{A}$ to any vector w in the space.

2. On the strategies of the two different proofs of Theorem 3.5.
The two different proofs of Theorem 3.5 given in this paper (with the second one excluding the case "$s + 3 = d_{\min}(A)$"; see below) follow two different strategies.

The first proof, given in Section 5, is based on vectors of grade $s + 2$, and "works its way up" to vectors of full grade $d_{\min}(A)$. This general strategy is similar to the one in the original paper of Faber and Manteuffel [4]. The details of our proof here, however, are quite different from [4]. In particular, simple arguments about the number of roots of certain polynomials (particularly in Lemmas 4.1 and 4.3) have replaced the continuity and topology arguments in the proof of [4]. We therefore consider this a simpler proof than the one given in [4].

The second proof, given in Section 6 works immediately with vectors of full grade $d_{\min}(A)$. We consider this approach more elementary than our first proof. We assume that the assertion of Theorem 3.5 is false, i.e. that A admits an optimal $(s+2)$-term recurrence, but is not normal(s). We show that if A is not normal(s), there must exist at least one initial vector v_1 of full grade $d = d_{\min}(A)$, for which the corresponding matrix H_d has a nonzero entry above its sth superdiagonal. If this nonzero entry already is in $H_{d,d-1}$, we are done. However, we cannot guarantee this, and therefore we need the Rotation Lemma to "rotate" a nonzero from the d-th column of H_d into the $(d-1)$-st column. This shows that A cannot admit an optimal $(s+2)$-term recurrence, contradicting our initial assumption.

3. The Rotation Lemma and the "missing case" $s + 3 = d_{\min}(A)$.
In the Rotation Lemma we "rotate" the nonzero entry $h_{1,d}$, where $d = d_{\min}(A)$, to give $\widetilde{h}_{1,d-1} \neq 0$ *or* $\widetilde{h}_{2,d-1} \neq 0$, cf. Fig 6.1. Therefore, the matrix $\widetilde{H}_{d,d-1}$ is *at least* $(d-1)$-band Hessenberg. The shortest possible optimal recurrence that A may admit hence is of length $d - 1$, or $s + 2$ for $s = d - 3$. The assumption that A admits an optimal recurrence of length $s + 3 < d_{\min}(A)$ then leads to a contradiction.

To prove also the "missing case" $s + 3 = d_{\min}(A)$, we need to guarantee that there exists a choice of s_1 so that $\widetilde{h}_{1,d-1} \neq 0$, giving a d-band Hessenberg matrix $\widetilde{H}_{d,d-1}$. Since Theorem 3.5 also holds for the case $s+3 = d_{\min}(A)$, we know that such s_1 *must exist*, but we were unable to prove the existence without using Theorem 3.5. Note, however, that in practical applications we are interested in recurrences of length

$s+2 \ll d_{\min}(A)$. Therefore the "missing case" of the Rotation Lemma is only of rather theoretical interest.

We point out that the construction given in the Rotation Lemma, namely the "structure-preserving" unitary transformation of an upper Hessenberg matrix, may be of interest beyond its application in our current context. To state this idea in a more general way, we introduce some notation. Let Ω_d be the set of the $d \times d$ unreduced upper Hessenberg matrices, and let $\Omega_d(s+2)$ be the subset consisting of the $(s+2)$-band Hessenberg matrices (these are unreduced by assumption; cf. Definition 3.1). Consider a *fixed* $H \in \Omega_d$, and define the set

$$\mathcal{R}_H \equiv \{\, G^* H G \in \Omega_d \;:\; G \text{ is unitary} \,\}.$$

Hence $\mathcal{R}_H$ is the set of all unitary transformations of H that are unreduced upper Hessenberg. Note that since $H \in \mathcal{R}_H$, the set $\mathcal{R}_H$ is nonempty. Using the Rotation Lemma (for $s+3 < d$) and Theorem 3.5 (for $s+3 = d$) the following result can be proven.

THEOREM 7.2. *Let s, d be given nonnegative integers, $s+2 < d$. For any $H \in \Omega_d$, the following assertions are equivalent:*

(1) H is I-normal(s),
i.e. $H^* = p(H)$ for a polynomial of (smallest possible) degree s.
(2) $\mathcal{R}_H \subset \Omega_d(s+2)$.

This result means that an unreduced upper Hessenberg matrix H is I-normal(s) if and only if H is $(s+2)$-band Hessenberg, and all unitary transformations that preserve the unreduced upper Hessenberg structure of H also preserve the $(s+2)$-band structure of H.

4. What distinguishes Theorem 3.5 from other results about normal operators.
Theorem 3.5 gives a *necessary* condition when an operator A is normal (of some degree s). This condition is also *sufficient*, as shown by Theorem 3.4. Hence this condition might be taken as a *definition* of normality, and it might be included among the numerous equivalent definitions in [8, 3]. We believe, however, that the nature of the result distinguishes it from the many other equivalent ones. This distinction is clear from the second proof given in Section 6.

Consider the linear operator A, and any cyclic invariant subspace $\mathcal{K}_d(A, v_1)$. Then the matrix representation of A with respect to the orthogonal basis $v_1, \ldots, v_d$ of $\mathcal{K}_d(A, v_1)$ generated by (3.1)–(3.3) is a $d \times d$ unreduced upper Hessenberg matrix H_d (cf. (6.2), where this is shown for the restriction of A to $\mathcal{K}_d(A, v_1)$). Typically, equivalent results for normality are derived using knowledge of *the whole matrix*, H_d in this case. But Theorem 3.5 is based only on knowledge of *a part of the matrix*, namely, the first $d-1$ columns of H_d. Our experience in this area shows that this difference also is the reason why Theorem 3.5 is rather difficult to prove, particularly when compared with other results about normal matrices or operators.

Acknowledgement. We thank Tom Manteuffel and an anonymous referee for suggestions that helped us to improve the presentation of the results.

REFERENCES

[1] W. E. ARNOLDI, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.

[2] E. W. Cheney, *Introduction to approximation theory*, McGraw-Hill Book Co., New York, 1966.
[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] D. K. Faddeev and V. N. Faddeeva, *Computational methods of linear algebra*, W. H. Freeman and Co., San Francisco, 1963.
[6] F. R. Gantmacher, *The theory of matrices. Vols. 1, 2*, Chelsea Publishing Co., New York, 1959.
[7] A. Greenbaum, *Iterative methods for solving linear systems*, vol. 17 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
[8] R. Grone, C. R. Johnson, E. M. de Sá, and H. Wolkowicz, *Normal matrices*, Linear Algebra and its Applications, 87 (1987), pp. 213–225.
[9] K. Hoffman and R. Kunze, *Linear Algebra*, Prentice-Hall, Englewood Cliffs, NJ, second ed., 1971.
[10] A. Kleppner, *The cyclic decomposition theorem*, Integral Equations Operator Theory, 25 (1996), pp. 490–495.
[11] J. Liesen and P. E. Saylor, *Orthogonal Hessenberg reduction and orthogonal Krylov subspace bases*, SIAM J. Numer. Anal., 42 (2005), pp. 2148–2158 (electronic).
[12] J. Liesen and Z. Strakoš, *On optimal short recurrences for generating orthogonal Krylov subspace bases*, SIAM Rev., (to appear).
[13] Y. Saad, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, second ed., 2003.