ON ORTHOGONAL REDUCTION TO HESSENBERG FORM WITH SMALL BANDWIDTH

V. FABER[*], J. LIESEN[†], AND P. TICHÝ[‡]

Abstract. Numerous algorithms in numerical linear algebra are based on the reduction of a given matrix A to a more convenient form. One of the most useful types of such reduction is the orthogonal reduction to (upper) Hessenberg form. This reduction can be computed by the Arnoldi algorithm. When A is Hermitian, the resulting upper Hessenberg matrix is tridiagonal, which is a significant computational advantage. In this paper we study necessary and sufficient conditions on A so that the orthogonal Hessenberg reduction yields a Hessenberg matrix with small bandwidth. This includes the orthogonal reduction to tridiagonal form as a special case. Orthogonality here is meant with respect to some given but unspecified inner product. While the main result is already implied by the Faber-Manteuffel theorem on short recurrences for orthogonalizing Krylov sequences, we consider it useful to present a new, less technical proof. Our proof utilizes the idea of a "minimal counterexample", which is standard in combinatorial optimization, but rarely used in the context of linear algebra.

Key words. reduction to Hessenberg form, Krylov subspace methods, Arnoldi method, Lanczos method

AMS subject classifications. 65F10, 65F25

1. Introduction. Many applications in engineering and science lead to linear algebraic problems involving a very large matrix $A \in \mathbb{C}^{N \times N}$. A common approach to solve such problems is to reduce A to a matrix that requires significantly less storage, or that is well suited for further processing. Algebraically, such reduction amounts to finding a more convenient basis for representing A.

One of the most useful types of such reduction is the *orthogonal reduction to (upper) Hessenberg form*, which is used, for example, in modern implementations of the QR method for solving eigenvalue problems (see [10] for a recent survey), and in the GMRES method for solving linear algebraic systems [9]. A standard method for computing this reduction is the Arnoldi algorithm [1]. Given a matrix A, an HPD matrix $B \in \mathbb{C}^{N \times N}$ defining the B-inner product $\langle x, y \rangle_B \equiv y^* B x$, and an initial vector $v_1 \in \mathbb{C}^N$, the Arnoldi algorithm generates a B-orthogonal basis for the (maximal) Krylov subspace of A and v_1.

More precisely, let d be the grade of v_1 with respect to A, i.e., the smallest possible degree of a polynomial p that satisfies $p(A)v_1 = 0$. Then the Arnoldi algorithm sequentially generates vectors $v_1, \ldots, v_d$, such that

$$(1.1) \quad \operatorname{span}\{v_1, \ldots, v_n\} = \operatorname{span}\{v_1, \ldots, A^{n-1}v_1\} \equiv \mathcal{K}_n(A, v_1), \quad n = 1, \ldots, d,$$

$$(1.2) \quad \langle v_i, v_j \rangle_B = 0, \qquad i \neq j, \quad i, j = 1, \ldots, d.$$

This is achieved by the following steps: For $n = 1, 2, \ldots,$

$$v_{n+1} = Av_n - \sum_{m=1}^{n} h_{m,n} v_m, \quad \text{where}$$

$$h_{m,n} = \frac{\langle Av_n, v_m \rangle_B}{\langle v_m, v_m \rangle_B}, \quad \text{for } m = 1, \dots, n.$$

If $v_{n+1} = 0$ then stop.

Here we have stated the classical Gram-Schmidt variant of the Arnoldi algorithm. For notational convenience, the basis vectors are not normalized. Other implementations are often preferable from a numerical point of view. In this paper, however, we assume exact arithmetic only, and do not consider differences in the finite precision behavior of different implementations.

Collecting the basis vectors in a matrix V_n, and the recurrence coefficients $h_{i,j}$ in a matrix H_n, the Arnoldi algorithm can be written in the following matrix form,

$$(1.3) \qquad AV_n = V_n H_n + v_{n+1} e_n^T, \quad n = 1, \dots, d.$$

Here e_n is the nth column of the identity matrix I_n and H_n is an $n \times n$ unreduced upper Hessenberg matrix given by

$$(1.4) \qquad H_n = \begin{bmatrix} h_{11} & \cdots & h_{1,n-1} & h_{1,n} \\ 1 & \ddots & \vdots & \vdots \\ & \ddots & h_{n-1,n-1} & h_{n-1,n} \\ & & 1 & h_{n,n} \end{bmatrix}.$$

The B-orthogonality of the basis vectors means that $V_n^* B V_n$ is an invertible $n \times n$ diagonal matrix, $n = 1, \dots, d$.

If d is the grade of v_1 with respect to A, then $\mathcal{K}_d(A, v_1)$ is A-invariant, and v_{d+1} must be the zero vector, so that the Arnoldi algorithm terminates at step $n = d$. Hence, at the dth iteration step the relation (1.3) becomes

$$(1.5) \qquad AV_d = V_d H_d.$$

Here H_d can be interpreted as the matrix representation of the linear operator A restricted to the A-invariant subspace $\mathcal{K}_d(A, v_1)$. Or, H_d can be interpreted as a *reduction of A to upper Hessenberg form*. For more about the theory and different implementations of the Arnoldi algorithm, we refer to [8, Chapter 6.3].

Now suppose that A is self-adjoint with respect to the B-inner product, i.e. that $\langle Ax, y \rangle_B = \langle x, Ay \rangle_B$ for all vectors $x, y \in \mathbb{C}^N$. This holds if and only if $A^* B = BA$, or, equivalently, the B-adjoint $A^+ \equiv B^{-1} A^* B$ satisfies $A^+ = A$. Denote $D \equiv V_d^* B V_d$. Since $A^+ = A$ we obtain, cf. (1.5),

$$H_d^* D = H_d^* V_d^* B V_d = V_d^* A^* B V_d = V_d^* B A^+ V_d = V_d^* B A V_d = D H_d.$$

Since D is diagonal, the upper Hessenberg matrix H_d must be tridiagonal. In other words, when A is self-adjoint with respect to the B-inner product, the Arnoldi algorithm B-orthogonally reduces A to tridiagonal form. In the special case $B = I$ and thus $A^+ = A^*$, the algorithm for computing this reduction is known as the Hermitian Lanczos algorithm [5].

Obviously, a reduction to tridiagonal form is very convenient from a numerical point of view. It is therefore of great interest to study necessary and sufficient conditions on A so that there exists an HPD matrix B for which A can be orthogonally

reduced to tridiagonal form or, more generally, to banded upper Hessenberg form with small bandwidth. Apart from trivial cases, the main necessary and sufficient condition on A is that there exists an HPD matrix B for which the B-adjoint A^+ is a low degree polynomial in A. As described in [7], this result is implied by the Faber-Manteuffel theorem on the existence of short recurrences for generating orthogonal Krylov subspace bases (in particular, see [7, Fig. 2.2]). Therefore, the question whether a given matrix is orthogonally reducible to banded Hessenberg form with low bandwidth has been completely answered. A separate proof of this result has been attempted in [6], but, as described in [7], that proof is based on less rigorous definitions and applies to nonderogatory matrices A only.

The purpose of this paper is to give a new proof of the necessary and sufficient conditions for orthogonal reducibility to upper Hessenberg form with small bandwidth. After recalling the sufficiency result from [7, Theorem 2.13], we first prove the necessity result for nonderogatory matrices (similarly as in [6], but starting from more rigorous definitions). We then show the general case inductively using a "minimal counterexample" argument. This is a standard argument in combinatorial optimization, but we have rarely seen this idea applied in linear algebra. Here we show that the smallest matrix giving a counterexample for the general case must be nonderogatory. Since we know from the first step of the proof that the result holds for nonderogatory matrices, no counterexample can possibly exist.

Reducibility to banded Hessenberg form with small bandwidth, particularly tridiagonal form, is a key property in many applications. Nevertheless, we are not aware that any complete proof of the necessary and sufficient conditions, that is independent of the technically more complicated result of Faber and Manteuffel, has appeared in the literature before. We point out that unlike the proofs of the Faber-Manteuffel theorem in [3, 2] (also cf. [11] for a related proof), our proof here is entirely based on linear algebra arguments. Furthermore, we believe that the general idea of our proof is of interest in its own right, which is a main reason for writing this paper.

2. Main definitions and sufficient conditions. Suppose that $A \in \mathbb{C}^{N \times N}$ is a given matrix, $B \in \mathbb{C}^{N \times N}$ is a given HPD matrix, and $v_1 \in \mathbb{C}^N$ is a given initial vector. (When A is real, we only consider real HPD matrices B and real initial vectors v_1.) We denote the degree of the minimal polynomial of A by $d_{\min}(A)$.

Consider the corresponding Hessenberg reduction of A as in (1.5). The Krylov subspace basis vectors $v_1, \ldots, v_d$ in this reduction are defined uniquely up to scaling by the conditions (1.1)–(1.2). This means that any other set of basis vectors $\widehat{v}_1, \ldots, \widehat{v}_d$ that also satisfies (1.1)–(1.2) is given by $\widehat{v}_n = \sigma_n v_n$ for some (nonzero) scalars $\sigma_1, \ldots, \sigma_n$. In matrix form, this can be written as $\widehat{V}_d = V_d S_d$, where $S_d = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$. Hence for this other basis, A satisfies the identity

$$A\widehat{V}_d = \widehat{V}_d \widehat{H}_d,$$

where $\widehat{H}_d = S_d^{-1} H_d S_d$. Clearly, the nonzero patterns of H_d and $\widehat{H}_d$ coincide. In particular, the upper bandwidth of H_d is independent of the algorithm that is used to compute the orthogonal reduction to Hessenberg form.

In this paper we are mostly interested in this upper bandwidth. We say that H_d is $(s+2)$-*band Hessenberg*, when the s-th superdiagonal of H_d contains at least one nonzero entry, and all entries above the s-th superdiagonal are zero. We can now rigorously define the concept of reducibility to banded upper Hessenberg form. We use the same definition as in [7, Definition 2.11], with the exception that here we do not require A to be nonsingular.

DEFINITION 2.1. *Let $A \in \mathbb{C}^{N \times N}$, let $B \in \mathbb{C}^{N \times N}$ be an HPD matrix, and let s be a nonnegative integer, $s + 2 \leq d_{\min}(A)$.*

(1) *If for an initial vector v_1 the matrix H_d in (1.5) is $(s + 2)$-band Hessenberg, then we say that A is reducible for the given B and v_1 to $(s + 2)$-band Hessenberg form.*

(2) *If A is reducible for the given B and any initial vector v_1 to at most $(s+2)$-band Hessenberg form, while it is reducible for the given B and at least one v_1 to $(s+2)$-band Hessenberg form, then we say that A is reducible for the given B to $(s + 2)$-band Hessenberg form.*

Note that by this definition the integer s is *uniquely determined*. This means that when A is reducible for the given B to $(s + 2)$-band Hessenberg form, then A is *not* reducible for this B to $(t + 2)$-band Hessenberg form for any $t \neq s$.

DEFINITION 2.2. *Let $A \in \mathbb{C}^{N \times N}$, and let $B \in \mathbb{C}^{N \times N}$ be HPD. Suppose that*

$$(2.1) \qquad A^+ \equiv B^{-1}A^*B = p_s(A),$$

where p_s is a polynomial of the smallest possible degree s having this property. Then A is called normal of degree s with respect to B, or, shortly, B-normal(s).

Using this definition, it is possible to prove the following sufficiency result for reducibility to $(s+2)$-band Hessenberg form; see [7, Theorem 2.13] (also cf. [4] for an analysis of the sufficient conditions in case $B = I$).

THEOREM 2.3. *Let $A \in \mathbb{C}^{N \times N}$, let $B \in \mathbb{C}^{N \times N}$ be an HPD matrix, and let s be a nonnegative integer, $s + 2 \leq d_{\min}(A)$. If A is B-normal(s), then A is reducible for the given B to $(s + 2)$-band Hessenberg form.*

Our statement of the sufficiency result is a little bit different from the one in [7, Theorem 2.13]. Here we assume that $s + 2 \leq d_{\min}(A)$, while [7, Theorem 2.13] assumes $s + 2 < d_{\min}(A)$. The assumption in [7] is made for notational consistency in that paper; extending the result to the case $s + 2 = d_{\min}(A)$ is straightforward. Furthermore, we have formulated the result for general matrices A, while in [7] it is assumed that A is nonsingular. The extension to the singular case is easy.

3. Necessary conditions. In this section we prove the reverse direction of Theorem 2.3, i.e., we show that if is reducible for the given B to $(s + 2)$-band Hessenberg form, where $s + 2 \leq d_{\min}(A)$, then A is B-normal(s). Our proof is based on three technical lemmas.

In the first lemma, we adopt [2, Lemma 4.3] to the notation used in this paper, and we generalize the assertion to include the case of singular A.

LEMMA 3.1. *Let $A \in \mathbb{C}^{N \times N}$, let $B \in \mathbb{C}^{N \times N}$ be an HPD matrix, and let s be a nonnegative integer, $s + 2 \leq d_{\min}(A)$. The matrix A is B-normal(s) if and only if,*

$$(3.1) \qquad A^+ v \in \mathcal{K}_{s+1}(A, v) \quad \text{for all vectors } v \text{ of grade } d_{\min}(A),$$

and there exists a vector v such that $A^+ v \notin \mathcal{K}_s(A, v)$.

Proof. Let A be B-normal(s). Then for each v, $A^+ v = p_s(A)v \in \mathcal{K}_{s+1}(A, v)$. Moreover, since s is the smallest degree of a polynomial for which $A^+ = p_s(A)$, there must exist a vector v such that $A^+ v \notin \mathcal{K}_s(A, v)$.

In the proof of the other direction we first suppose that A is nonsingular. Then by [2, Lemma 4.3], (3.1) implies that A is B-normal(t) for some $t \leq s$. Since there exists a vector v such that $A^+ v \notin \mathcal{K}_s(A, v)$, we must have $t \geq s$, and thus $t = s$.

Now suppose that A is singular. Then there exists a scalar $\mu \in \mathbb{C}$ such that $C \equiv A + \mu I$ is nonsingular. Clearly, $d_{\min}(A) = d_{\min}(C)$. Furthermore, note that for any vector v of grade $d_{\min}(A)$, we have $\mathcal{K}_{s+1}(A, v) = \mathcal{K}_{s+1}(C, v)$. Moreover, since

$$A^+ = B^{-1}A^*B = B^{-1}C^*B - \overline{\mu}I = C^+ - \overline{\mu}I,$$

$A^+v \in \mathcal{K}_{s+1}(A, v)$ holds if and only if $C^+v \in \mathcal{K}_{s+1}(C, v)$. Hence, if the singular matrix A satisfies the assertion, then the nonsingular matrix $C = A + \mu I$ satisfies the assertion as well, so that C must be B-normal(s). But $C^+ = p_s(C)$ implies that $A^+ = q_s(A)$, where q_s is a polynomial of (smallest possible) degree s. Hence A is B-normal(s) as well, which finishes the proof. $\square$

In the next lemma we prove the necessity result for nonderogatory matrices A (see also [6, pp. 2156–2157] for a similar argument).

LEMMA 3.2. *Let $A \in \mathbb{C}^{N \times N}$ be a nonderogatory matrix, i.e., $d_{\min}(A) = N$. Let $B \in \mathbb{C}^{N \times N}$ be an HPD matrix, and let s be a nonnegative integer, $s + 2 \leq d_{\min}(A)$. If A is reducible for the given B to $(s + 2)$-band Hessenberg form, then A is B-normal(s).*

Proof. We prove the assertion by contradiction. Suppose that A is reducible for the given B to $(s + 2)$-band Hessenberg form, but that A is *not* B-normal(s). By Lemma 3.1, there either exists an integer $t < s$ such that $A^+v_1 \in \mathcal{K}_{t+1}(A, v_1)$ for all vectors v_1, or there exists a vector v_1 of grade $d_{\min}(A) = N$ such that $A^+v_1 \notin \mathcal{K}_{s+1}(A, v_1)$.

In the first case, one can easily show that the matrix A is reducible to (at most) $(t + 2)$-band Hessenberg form, which is a contradiction since $t < s$.

In the second case, consider a vector v of grade N such that $A^+v_1 \notin \mathcal{K}_{s+1}(A, v_1)$. Since v_1 is of full grade, we know that there exist scalars $\beta_1, \ldots, \beta_N \in \mathbb{C}$, such that

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

By assumption, at least one β_j, $s + 2 \leq j \leq N$, is nonzero. If this nonzero scalar is β_k, then the entry $h_{1,k}$ of H_N satisfies

$$h_{1,k} = \frac{\langle Av_k, v_1 \rangle_B}{\langle v_1, v_1 \rangle_B} = \frac{\langle v_k, A^+v_1 \rangle_B}{\langle v_1, v_1 \rangle_B} = \overline{\beta}_k \frac{\langle v_k, v_k \rangle_B}{\langle v_1, v_1 \rangle_B} \neq 0.$$

But since $k \geq s + 2$, this means that H_d is *not* $(s + 2)$-band Hessenberg, which contradicts our assumption that A is reducible to $(s + 2)$-band Hessenberg form. $\square$

We next show that the "minimal counterexample" of a matrix A that is reducible for the given B to $(s+2)$-band Hessenberg form but that is *not* B-normal(s) must be nonderogatory.

LEMMA 3.3. *Suppose that s is a given nonnegative integer. Let A be a square matrix of smallest possible dimension N and with $d_{\min}(A) \geq s + 2$ such that the following holds: There exists HPD matrix $B \in \mathbb{C}^{N \times N}$ such that*
1. A is reducible for the given B to $(s + 2)$-band Hessenberg form,
2. A is not B-normal(s).
Then A is nonderogatory (i.e. $d_{\min}(A) = N$).

Proof. Suppose that A is a matrix that satisfies the assumptions, and that B is the corresponding HPD matrix for which A is reducible to $(s + 2)$-band Hessenberg form,

but with respect to which A is not normal of degree s. Let the Jordan normal form of A be given by $A = WJW^{-1}$, where $J = J_1 \oplus \cdots \oplus J_k$ with eigenvalues $\lambda_1, \ldots, \lambda_k$, and corresponding invariant subspaces of dimensions $s_1, \ldots, s_k$, respectively. If $k = 1$, then A is nonderogatory and we are done. Hence we may assume that $k > 1$.

Suppose that v_1 is any initial vector of grade d with respect to A and consider the corresponding Hessenberg reduction (1.5) using the B-inner product. Using the Jordan normal form of A, it is easy to see that this Hessenberg reduction is equivalent with

$$(3.2) \qquad J\widehat{V}_d = \widehat{V}_d H_d, \qquad \widehat{V}_d^* \widehat{B} \widehat{V}_d \quad \text{diagonal},$$

where $\widehat{V}_d \equiv W^{-1} V_d$ and $\widehat{B} \equiv W^* B W$, which is HPD. Note that the Hessenberg matrices in the Hessenberg reduction of A and in (3.2) coincide. Since A is reducible for the given B to $(s+2)$-band Hessenberg form, J is reducible for the given $\widehat{B}$ to $(s+2)$-band Hessenberg form (and vice versa).

It suffices to show that J is nonderogatory. Suppose not. Then there are two Jordan blocks, say J_1 and J_2 with $s_1 \leq s_2$, that correspond to the same eigenvalue (i.e. $\lambda_1 = \lambda_2$). Define the $(N - s_1) \times (N - s_1)$ matrix $\widetilde{J} \equiv J_2 \oplus \cdots \oplus J_k$, which satisfies $d_{\min}(\widetilde{J}) = d_{\min}(J) \geq s + 2$. Now define an inner product $[\cdot, \cdot]$ on $\mathbb{C}^{N-s_1} \times \mathbb{C}^{N-s_1}$ by

$$(3.3) \qquad [x, y] \equiv \langle 0_{s_1} \oplus x, 0_{s_1} \oplus y \rangle_{\widehat{B}}.$$

Here 0_{s_1} denotes the zero vector of length s_1. This inner product is generated by an HPD matrix $\widetilde{B}$, $[x, y] = y^* \widetilde{B} x$ for all vectors x and y. Using the standard basis vectors and the definition of $[\cdot, \cdot]$ it is easy to show that $\widetilde{B}$ is the $(N - s_1) \times (N - s_1)$ trailing principal submatrix of $\widehat{B}$ (using MATLAB notation, $\widetilde{B} = \widehat{B}(1 + s_1 : N, 1 + s_1 : N)$).

If y_1 is any initial vector of grade d with respect to $\widetilde{J}$, then $v_1 = 0_{s_1} \oplus y_1$ is of grade d with respect to J. By construction, the corresponding Hessenberg reductions of $\widetilde{J}$ and J using the $\widetilde{B}$- and $\widehat{B}$-inner products, respectively, lead to the same unreduced upper Hessenberg matrix H_d. Consequently, the matrix $\widetilde{J}$ is reducible for $\widetilde{B}$ to $(s+2)$-band Hessenberg form.

Since $N - s_1 < N$, our initial assumption implies that the matrix $\widetilde{J}$ is $\widetilde{B}$-normal(s). Then [7, Theorem 3.1] shows: First, $\widetilde{J}$ is diagonalizable and hence diagonal, in particular $s_2 = 1$. Second, assuming that the eigenvalues of $\widetilde{J}$ are ordered so that the same eigenvalues form a single block, the HPD matrix $\widetilde{B}$ is block diagonal with block sizes corresponding to those of $\widetilde{J}$. Third, there exists a polynomial p_s of smallest possible degree s such that $p_s(\widetilde{J}) = \widetilde{J}^*$ (i.e., $p_s(\lambda_j) = \overline{\lambda}_j$ for all eigenvalues λ_j of A).

Consequently, J is diagonal with the first two eigenvalues equal, and $p_s(J) = J^*$, where p_s is a polynomial of smallest possible degree with this property. Moreover, $\widehat{B}$ is HPD and block diagonal with block sizes corresponding to those of J, except for possibly its first row and column. For simplicity of the presentation, we assume that $\widehat{B}$ is diagonal except for its first row and column; the argument for the *block* diagonal case is more technical but mathematically analogous. Then $\widehat{B}$ has the nonzero structure

$$\widehat{B} = \begin{bmatrix} \star & \star & \cdots & \star \\ \star & \star & & \\ \vdots & & \ddots & \\ \star & & & \star \end{bmatrix}.$$

Now we reverse the roles of J_1 and J_2 and repeat the whole construction. More specifically, we denote the columns of the matrix W (from the Jordan decomposition of A)

by $w_1, \ldots, w_N$. Then $A = WJW^{-1} = W_1 J W_1^{-1}$, where $W_1 \equiv [w_2, w_1, w_3, \ldots, w_N]$. Here we have used that $J_1 = J_2$ and that J is diagonal. Repeating the above construction yields a matrix $B_1 = W_1^* B W_1$, which is of the same form as $\widehat{B}$, i.e.

$$B_1 = W_1^* B W_1 = \begin{bmatrix} \star & \star & \cdots & \star \\ \star & \star & & \\ \vdots & & \ddots & \\ \star & & & \star \end{bmatrix} .$$

In particular, by comparing the second row on both sides of this equation, we see that

$$w_1^* B [w_2, w_1, w_3, \ldots, w_N] = [\star, \star, 0, \cdots, 0] .$$

Then the first row of $\widehat{B}$ is given by $w_1^* B W = [\star, \star, 0, \cdots, 0]$, which shows that indeed $\widehat{B}$ is block diagonal with block sizes corresponding to those of J. Hence the $N \times N$ matrix J is $\widehat{B}$-normal(s), which contradicts our assumption and completes the proof. □

In the following theorem we state the main result of this paper. The sufficiency part has already been stated in Theorem 2.3 above and is repeated here for completeness.

THEOREM 3.4. *Let* $A \in \mathbb{C}^{N \times N}$, *let* $B \in \mathbb{C}^{N \times N}$ *be an HPD matrix, and let s be a nonnegative integer,* $s + 2 \leq d_{\min}(A)$. *The matrix A is B-normal(s) if and only if A is reducible for the given B to $(s + 2)$-band Hessenberg form.*

Proof. We only have to show that if A is reducible for the given B to $(s+2)$-band Hessenberg form, then A is B-normal(s). By Lemma 3.2, this statement is true for nonderogatory matrices A. However, by Lemma 3.3, the minimal counterexample is nonderogatory. Hence there is no minimal counterexample, so that the assertion must hold. □

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] V. FABER, J. LIESEN, AND P. TICHÝ, *The Faber-Manteuffel theorem for linear operators*, SIAM J. Numer. Anal., 46 (2008), pp. 1323–1337.
[3] 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.
[4] T. HUCKLE, *The Arnoldi method for normal matrices*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 479–489.
[5] C. LANCZOS, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards, 45 (1950), pp. 255–282.
[6] J. LIESEN AND P. E. SAYLOR, *Orthogonal Hessenberg reduction and orthogonal Krylov subspace bases*, SIAM J. Numer. Anal., 42 (2005), pp. 2148–2158.
[7] J. LIESEN AND Z. STRAKOŠ, *On optimal short recurrences for generating orthogonal Krylov subspace bases*, SIAM Rev., (to appear).
[8] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, second ed., 2003.
[9] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
[10] D. S. WATKINS, *The QR algorithm revisited*, SIAM Rev., 50 (2008), pp. 133–145.
[11] H. ZHA AND Z. ZHANG, *The Arnoldi process, short recursions, and displacement ranks*, Linear Algebra Appl., 249 (1996), pp. 169–188.