Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Linear Algebra and its Applications 414 (2006) 435–444

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Necessary and sufficient conditions for orthogonal similarity transformations to obtain the Arnoli(Lanczos)–Ritz values ☆

Raf Vandebril *, Marc Van Barel

Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium

Received 12 July 2004; accepted 6 October 2005
Available online 20 December 2005
Submitted by V. Mehrmann

Abstract

It is a well-known fact that while reducing a symmetric matrix into a similar tridiagonal one, the already tridiagonal matrix in the partially reduced matrix has as eigenvalues the Lanczos–Ritz values. This behavior is also shared by the reduction algorithm which transforms symmetric matrices via orthogonal similarity transformations to semiseparable form. Moreover also the orthogonal reduction to Hessenberg form has a similar behavior with respect to the Arnoldi–Ritz values.

In this paper we investigate the orthogonal similarity transformations creating this behavior. Two easy conditions are derived, which provide necessary and sufficient conditions, such that the partially reduced matrices have the desired convergence behavior. The conditions are easy to check as they demand that in every step of the reduction algorithm two particular matrices need to have a zero block.

Keywords: Ritz values; Arnoldi–Ritz values; Lanczos–Ritz values; Similarity transformations

☆ The research was partially supported by the Research Council K.U. Leuven, project OT/00/16 (SLAP: Structured Linear Algebra Package), by the Fund for Scientific Research—Flanders (Belgium), projects G.0078.01 (SMA: Structured Matrices and their Applications), G.0176.02 (ANCILA: Asymptotic aNalysis of the Convergence behavior of Iterative methods in numerical Linear Algebra), G.0184.02 (CORFU: Constructive study of Orthogonal Functions) and G.0455.0 (RHPH: Riemann–Hilbert problems, random matrices and Padé–Hermite approximation), and by the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture, project IUAP V-22 (Dynamical Systems and Control: Computation, Identification and Modelling). The scientific responsibility rests with the authors.

* Corresponding author.
E-mail addresses: raf.vandebril@cs.kuleuven.be (R. Vandebril), marc.vanbarel@cs.kuleuven.be (M. Van Barel).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2005.10.026

1. Introduction

It is well-known that while reducing a symmetric matrix into a similar tridiagonal one, the intermediate tridiagonal matrices contain the Lanczos–Ritz values as eigenvalues. Or for a Hessenberg matrix they contain the so-called Arnoldi–Ritz values. More information can be found in the following books [1,2,4,6–8] and the references therein.

The goal of this paper is not to investigate the convergence behavior of the Ritz-values (see e.g. [5] and the references therein), nor to prove that certain matrices have close connections with the Lanczos(Arnoldi)–Ritz values (see e.g. [3,4]). Our goal is to investigate the orthogonal similarity transformations in general causing this behavior. To achieve this goal we assume that the performed similarity transformation leads to this special convergence behavior. This gives two necessary conditions which always have to be satisfied. Based on these two conditions we prove that orthogonal similarity transformations inheriting these conditions have the desired convergence behavior. In this way we derived necessary and sufficient conditions. Using the conditions, which are straightforward to check, it is an easy exercise to prove that the orthogonal similarity transformations of matrices to semiseparable, tridiagonal and/or Hessenberg form share the same convergence behavior, with respect to the Lanczos(Arnoldi)–Ritz values.

The convergence behavior of the Lanczos(Arnoldi)–Ritz values towards the eigenvalues [5] plays an important role in several applications where one wants to locate specific parts of the spectrum. Recently the orthogonal similarity transformation of a symmetric matrix into a similar semiseparable one was derived [9]. This reduction combines the Lanczos–Ritz values behavior, together with a nested subspace iteration. The knowledge that the subspace iteration works on the Ritz values is crucial for understanding the convergence behavior in the reduction algorithm. This combined convergence behavior can also be found in the reduction to a similar semiseparable plus diagonal matrix [11] or the reduction towards upper triangular semiseparable or Hessenberg-like form [10]. For these reductions it is important to know that the Ritz values appear in certain blocks in the matrix. This paper provides easy conditions to check whether these reduction algorithms and others, e.g. the reduction to tridiagonal or Hessenberg matrices, have the Lanczos(Arnoldi)–Ritz values appearing in a submatrix.

The paper is organized as follows. In Section 2 we introduce briefly the type of orthogonal similarity transformation considered, and also the notion of Ritz values and Krylov subspaces is briefly refreshed. The orthogonal similarity transformations obeying the desired convergence behavior are investigated in Section 3. This leads to two simple, but necessary conditions. In Section 4 we prove that the conditions derived in the previous section are also sufficient to obtain that the partial reduced matrices have the Lanczos(Arnoldi)–Ritz values. The previous two sections describe the generic case, namely with one Krylov subspace. In case of an invariant Krylov subspace the theory changes slightly. In Section 5 we investigate, what happens in the case of invariant subspaces. Some general remarks, and an extra property of the orthogonal matrices are derived in Section 6. The final section of the paper contains the conclusions.

2. Ritz values and Arnoldi(Lanczos)–Ritz values

We will briefly introduce here the notion of “Ritz values”, related to the orthogonal similarity transformation. The orthogonal similarity transformations we consider are based on finite induction. In each induction step a row and a column are added to the desired structure. In this way all the columns and rows are transformed, such that the resulting matrix satisfies the desired structure. Suppose, we have a matrix $A^{(0)} = A$, which is transformed via an initial orthogonal similarity transformation into the matrix $A^{(1)} = Q_0^{\mathrm{T}} A^{(0)} Q_0$. The initial transformation Q_0 is not essential in the following proof, as it does not affect the reduction algorithms. It does have an effect on the convergence behavior of the reduction, as will be shown in this subsection. We remark that in real applications, often a matrix Q_0 is chosen in such a way to obtain a specific convergence behavior.

The other orthogonal transformations Q_k, $1 \leqslant k \leqslant n-1$, are constructed by the reduction algorithms. Let us denote the orthogonal transformation to go from $A^{(k)}$ to $A^{(k+1)}$ as Q_k, and we denote with $Q_{0:k}$ the orthogonal matrix equal to the product $Q_0 Q_1 \ldots Q_k$. This means that

$$\begin{aligned} A^{(k+1)} &= Q_k^{\mathrm{T}} A^{(k)} Q_k \\ &= Q_k^{\mathrm{T}} Q_{k-1}^{\mathrm{T}} \ldots Q_1^{\mathrm{T}} Q_0^{\mathrm{T}} A Q_0 Q_1 \ldots Q_{k-1} Q_k \\ &= Q_{0:k}^{\mathrm{T}} A Q_{0:k}. \end{aligned}$$

The matrix $A^{(k+1)}$ is of the following form:

$$\left(\begin{array}{c|c} R_{k+1} & \times \\ \hline \times & A_{k+1} \end{array}\right),$$

where R_{k+1} stands for that part of the matrix of dimension $(k+1) \times (k+1)$ which is already transformed to the appropriate form, e.g. tridiagonal, semiseparable, Hessenberg, etc. The matrix A_{k+1} is of dimension $(n-k-1) \times (n-k-1)$. The $\times$ denote arbitrary matrices. They are unimportant in the remaining part of the exposition. Remark however that the matrices $A^{(k)}$ are not necessarily symmetric, as the elements $\times$ may falsely indicate.

Let us partition the matrix $Q_{0:k}$ as follows:

$$Q_{0:k} = \left(\overleftarrow{Q}_{0:k} | \overrightarrow{Q}_{0:k}\right) \quad \text{with} \begin{cases} \overleftarrow{Q}_{0:k} \in \mathbb{R}^{n \times (k+1)}, \\ \overrightarrow{Q}_{0:k} \in \mathbb{R}^{n \times (n-k-1)}. \end{cases}$$

This means,

$$A \left(\overleftarrow{Q}_{0:k} | \overrightarrow{Q}_{0:k}\right) = \left(\overleftarrow{Q}_{0:k} | \overrightarrow{Q}_{0:k}\right) \left(\begin{array}{c|c} R_{k+1} & \times \\ \hline \times & A_{k+1} \end{array}\right).$$

The eigenvalues of R_{k+1} are called the Ritz values of A with respect to the subspace spanned by the columns of $\overleftarrow{Q}_{0:k}$ (see e.g. [2]).

Suppose we have now the Krylov subspace of order k with initial vector v:

$$\mathscr{K}_k(A, v) = \langle v, Av, \ldots, A^{k-1} v \rangle,$$

where $\langle x, y, z \rangle$ denotes the vector space spanned by the vectors x, y and z. For simplicity we assume in Sections 3 and 4 that the Krylov subspaces we are working with are not invariant, i.e. that for every k: $\mathscr{K}_k(A, v) \neq \mathscr{K}_{k+1}(A, v)$, where $k = 1, 2, \ldots, n-1$. The special case of invariant subspaces is dealt with in Section 5.

If the columns of the matrix $\overleftarrow{Q}_{0:k}$ form an orthonormal basis of the Krylov subspace $\mathscr{K}_{k+1}(A, v)$, then we say that the eigenvalues of R_{k+1} are called the Arnoldi–Ritz values of A with respect to the initial vector v. If the matrix A is symmetric, one often calls the Ritz values the Lanczos–Ritz values.

3. Necessary conditions to obtain the Arnoldi(Lanczos)–Ritz values as eigenvalues in the already reduced block of the matrix

In this section, we investigate the properties of orthogonal similarity transformations, where the eigenvalues in the already reduced block of the matrix are the Arnoldi–Ritz values, with respect to the starting vector v, where $v/\|v\| = \pm Q_0 e_1$. This makes clear that the initial transformation can change the convergence behavior, as it changes the Krylov subspace and hence also the Ritz values. We remark once more that this initial transformation does not change the reduction algorithm as the actual algorithm reduces the matrix $A^{(1)} = Q_0^{\mathrm{T}} A Q_0$ to the desired form. However, in practice a good choice of the vector v, can have important consequences for the convergence behavior in applications.

Suppose that our orthogonal similarity reduction of the matrix into another matrix has the following form after step $k-1$ (with $k = 1, 2, \ldots, n-1$):

$$\begin{pmatrix} R_k & \times \\ \times & \times \end{pmatrix} = Q_{0:k-1}^{\mathrm{T}} A Q_{0:k-1}.$$

This means that we start with this matrix at step k of the reduction: with R_k a square matrix of dimension k, which has as eigenvalues the Arnoldi–Ritz values. Hence, we have the following properties for the orthogonal matrix $Q_{0:k-1}$:

(1) The columns of $\overleftarrow{Q}_{0:k-1}$ form an orthogonal basis for $\mathscr{K}_k(A, v)$.
(2) The columns of $\overrightarrow{Q}_{0:k-1}$ form an orthogonal basis for the orthogonal complement of $\mathscr{K}_k(A, v)$.

As already mentioned before, for simplicity reasons we assume here that we work with one non-invariant Krylov subspace $\mathscr{K}_k(A, v)$. The more general case is dealt with in Section 5.

After the next step in the transformation we have that the block R_{k+1} has the Ritz values as eigenvalues with respect to $\mathscr{K}_{k+1}(A, v)$. This results in two easy conditions, similar to the ones described above. After step k, in the beginning of step $k+1$ we have:

(1) The columns of $\overleftarrow{Q}_{0:k}$ form an orthogonal basis for $\mathscr{K}_{k+1}(A, v) = \mathscr{K}_k(A, v) + \langle A^k v \rangle$.
(2) The columns of $\overrightarrow{Q}_{0:k}$ form an orthogonal basis for the orthogonal complement of $\mathscr{K}_{k+1}(A, v)$.

We have the following equalities:

$$\begin{aligned} A &= Q_{0:k-1} A^{(k)} Q_{0:k-1}^{\mathrm{T}} \\ &= Q_{0:k} A^{(k+1)} Q_{0:k}^{\mathrm{T}}. \end{aligned}$$

This means that the transformation to go from matrix $A^{(k)}$ to matrix $A^{(k+1)}$ can also be written in the following form:

$$Q_{0:k}^{\mathrm{T}} Q_{0:k-1} A^{(k)} Q_{0:k-1}^{\mathrm{T}} Q_{0:k} = A^{(k+1)}.$$

Using the fact that Q_k denotes the orthogonal matrix to go from matrix $A^{(k)}$ to matrix $A^{(k+1)}$, we get:

$$
\begin{aligned}
Q_k^{\mathrm{T}} &= Q_{0:k}^{\mathrm{T}} Q_{0:k-1} \\
&= \begin{pmatrix} \overleftarrow{Q}_{0:k}^{\mathrm{T}} \\ \overrightarrow{Q}_{0:k}^{\mathrm{T}} \end{pmatrix} \left(\overleftarrow{Q}_{0:k-1} \,\middle|\, \overrightarrow{Q}_{0:k-1} \right) \\
&= \begin{pmatrix} (Q_k)_{11}^{\mathrm{T}} & (Q_k)_{12}^{\mathrm{T}} \\ (Q_k)_{21}^{\mathrm{T}} & (Q_k)_{22}^{\mathrm{T}} \end{pmatrix},
\end{aligned}
$$

where the $(Q_k)_{11}^{\mathrm{T}}, (Q_k)_{12}^{\mathrm{T}}, (Q_k)_{21}^{\mathrm{T}}$ and $(Q_k)_{22}^{\mathrm{T}}$ denote a partitioning of the matrix Q_k^{T}. These blocks have the following dimensions: $(Q_k)_{11}^{\mathrm{T}} \in \mathbb{R}^{(k+1)\times k}$, $(Q_k)_{12}^{\mathrm{T}} \in \mathbb{R}^{(k+1)\times(n-k)}$, $(Q_k)_{21}^{\mathrm{T}} \in \mathbb{R}^{(n-k-1)\times k}$ and $(Q_k)_{22}^{\mathrm{T}} \in \mathbb{R}^{(n-k-1)\times(n-k)}$. It can be seen rather easily, by combining the properties of the matrices $Q_{0:k-1}$ and $Q_{0:k}$ from above, that the block $(Q_k)_{21}^{\mathrm{T}}$ has to be zero. This zero block in the matrix Q_k is the first necessary condition.

To obtain a second condition, we will investigate the structure of an intermediate matrix $\widetilde{A}^{(k)}$ satisfying

$$
\begin{aligned}
\widetilde{A}^{(k)} &= Q_k^{\mathrm{T}} A^{(k)} \\
&= Q_k^{\mathrm{T}} Q_{0:k-1}^{\mathrm{T}} A Q_{0:k-1} \\
&= Q_{0:k}^{\mathrm{T}} A Q_{0:k-1},
\end{aligned}
$$

which can be rewritten as:

$$Q_{0:k}\widetilde{A}^{(k)} = A Q_{0:k-1}. \tag{1}$$

Rewriting Eq. (1) gives us:

$$A\left(\overleftarrow{Q}_{0:k-1} | \overrightarrow{Q}_{0:k-1}\right) = \left(\overleftarrow{Q}_{0:k} | \overrightarrow{Q}_{0:k}\right)\widetilde{A}^{(k)}.$$

Because the columns of $A\overleftarrow{Q}_{0:k-1}$ belong to the Krylov subspace: $\mathscr{K}_{k+1}(A, v)$, which is spanned by the columns of $\overleftarrow{Q}_{0:k}$, we have that $\widetilde{A}^{(k)}$ has a zero block of dimension $(n-k-1)\times k$ in the lower left corner. This provides us a second condition.

The two conditions presented here, namely the condition on $\widetilde{A}^{(k)}$ and the condition on Q_k, are necessary to have the desired convergence properties in the reduction. In the next section we will prove that they are also sufficient. We will formulate this as a theorem:

Theorem 1. *Suppose, we apply an orthogonal similarity transformation on the matrix A (as described in Section 2), such that the already reduced part R_k in the matrix has the Arnoldi–Ritz values in each step of the reduction algorithm. Then we have the following two properties for every $1 \leqslant k \leqslant n-1$:*

- *The matrix Q_k^{T}, which is the orthogonal matrix to transform $A^{(k)}$ into the matrix $A^{(k+1)} = Q_k^{\mathrm{T}} A^{(k)} Q_k$ has a zero block of dimension $(n-k-1)\times k$ in the lower left corner.*
- *The matrix $\widetilde{A}^{(k)} = Q_k^{\mathrm{T}} A^{(k)}$ has a zero block of dimension $(n-k-1)\times k$ in the lower left corner.*

4. Sufficient conditions to obtain the convergence behavior

We prove that the properties from Theorem 1 connected to the matrices Q_k and $\widetilde{A}^{(k)}$ are sufficient to have the Arnoldi–Ritz values as eigenvalues in the blocks R_k.

Theorem 2. *Suppose, we apply an orthogonal similarity transformation on the matrix A (as described in Section 2), such that we have for $A^{(0)} = A$:*

$$Q_0 e_1 = \pm \frac{v}{\|v\|} \quad \text{and} \quad Q_0^{\mathrm{T}} A^{(0)} Q_0 = A^{(1)}.$$

Assume that the corresponding Krylov subspace $\mathscr{K}_k(A, v)$, will not become invariant for $k \leqslant n-1$. Suppose that for every step $1 \leqslant k \leqslant n-1$ of the reduction algorithm we have the following two properties:

- *the matrix Q_k^{T}, which is the orthogonal matrix to transform $A^{(k)}$ into the matrix $A^{(k+1)} = Q_k^{\mathrm{T}} A^{(k)} Q_k$ has a zero block of dimension $(n-k-1) \times k$ in the lower left corner;*
- *the matrix $\widetilde{A}^{(k)} = Q_k^{\mathrm{T}} A^{(k)}$ has a zero block of dimension $(n-k-1) \times k$ in the lower left corner.*

Then we have that for the matrix $A^{(k+1)}$ partitioned as

$$A^{(k+1)} = \left(\begin{array}{c|c} R_{k+1} & \times \\ \hline \times & A_{k+1} \end{array} \right),$$

the matrix R_{k+1} of dimension $(k+1) \times (k+1)$ has the Ritz values with respect to the Krylov space $\mathscr{K}_{k+1}(A, v)$.

Proof. We will prove the theorem by induction on k.

Step 1. The theorem is true for $k = 1$, because $Q_0^{\mathrm{T}} A Q_0$ contains clearly the Arnoldi–Ritz value in the upper left 1×1 block.

Step k. Suppose the theorem is true for $A^{(1)}, A^{(2)}, \ldots, A^{(k)}$, with $k \leqslant n-1$. This means that the columns of $\overleftarrow{Q}_{0:k-1}$ span the Krylov subspace $\mathscr{K}_k(A, v)$. Then we will prove now that the conditions are sufficient to have that the columns of $\overleftarrow{Q}_{0:k}$ span the Krylov subspace of $\mathscr{K}_{k+1}(A, v)$. We have the following equalities

$$\begin{aligned} \widetilde{A}^{(k)} &= Q_k^{\mathrm{T}} A^{(k)} \\ &= Q_k^{\mathrm{T}} Q_{0:k-1}^{\mathrm{T}} A Q_{0:k-1} \\ &= Q_{0:k}^{\mathrm{T}} A Q_{0:k-1}. \end{aligned}$$

Therefore,

$$\begin{aligned} A Q_{0:k-1} &= Q_{0:k} \widetilde{A}^{(k)} \\ A \left(\overleftarrow{Q}_{0:k-1} | \overrightarrow{Q}_{0:k-1} \right) &= \left(\overleftarrow{Q}_{0:k} | \overrightarrow{Q}_{0:k} \right) \widetilde{A}^{(k)}. \end{aligned}$$

Hence, we have already that the columns of $A \overleftarrow{Q}_{0:k-1}$ are part of the space spanned by the columns of $\overleftarrow{Q}_{0:k}$. Note that the columns of $A \overleftarrow{Q}_{0:k-1}$ span the same space as $A \mathscr{K}_k(A, v)$. We have the following relation:

$$A \mathscr{K}_k(A, v) \subseteq \mathrm{Range}(\overleftarrow{Q}_{0:k}). \tag{2}$$

With $\mathrm{Range}(A)$ we denote the vector space spanned by the columns of the matrix A. We also have that:

$$Q_{0:k} = Q_{0:k-1} Q_k,$$

$$Q_{0:k} Q_k^{\mathrm{T}} = Q_{0:k-1}.$$

Hence,

$$\left(\overleftarrow{Q}_{0:k} | \overrightarrow{Q}_{0:k}\right) Q_k^{\mathrm{T}} = \left(\overleftarrow{Q}_{0:k-1} | \overrightarrow{Q}_{0:k-1}\right).$$

Using the zero structure of the matrix Q_k^{T} we have:

$$\mathrm{Range}(\overleftarrow{Q}_{0:k-1}) = \mathscr{K}_k(A, v) \subseteq \mathrm{Range}(\overleftarrow{Q}_{0:k}).$$

When we combine this, with Eq. (2) and the fact that our subspace $\mathscr{K}_k(A, v)$ is not invariant, i.e. $\mathscr{K}_{k+1}(A, v) \neq \mathscr{K}_k(A, v)$, we get:

$$\mathrm{Range}(\overleftarrow{Q}_{0:k}) = \mathscr{K}_{k+1}(A, v).$$

This proves the theorem for $A^{(k+1)}$.

5. The case of invariant subspaces

The theorems and proofs of the previous sections were based on the fact that the Krylov subspace $\mathscr{K}_k(A, v)$ was never invariant. In the case of an invariant subspace, we can apply deflation and continue working on the deflated part, or we can derive similar theorems as in the previous sections, but for combined Krylov subspaces. We will investigate these two possibilities in deeper detail in this section.

5.1. Some notation

We will introduce some extra notation to be able to work in an efficient way with these combined Krylov subspaces. Let us denote with $K_k(A)$ the following subspace:

$$K_k(A) = \mathscr{K}_{l_1}(A, v_1) \cup \mathscr{K}_{l_2}(A, v_2) \cup \cdots \cup \mathscr{K}_{l_{t-1}}(A, v_{t-1}) \cup \mathscr{K}_{k_t}(A, v_t).$$

This means that $K_k(A)$ is the union of t different Krylov subspaces. We make the following assumptions:

- The dimension of $K_k(A)$ equals k, and $k = l_1 + l_2 + l_3 + \cdots + l_{t-1} + k_t$.
- With l_i we denote the maximum dimension the Krylov subspace with matrix A and initial vector v_i can reach, before becoming invariant. This means that the Krylov subspaces $\mathscr{K}_p(A, v_i)$ with $p \geqslant l_i$ are all equal to $\mathscr{K}_{l_i}(A, v_i) \neq \mathscr{K}_{l_i-1}(A, v_i)$. The k_i is an index $1 \leqslant k_i \leqslant l_i$ for the Krylov subspace with matrix A and initial vector v_i:

  $$\mathscr{K}_{k_i}(A, v_i) = \langle v_i, Av_i, \ldots, A^{k_i-1} v_i \rangle.$$

- The starting vectors v_i of the different Krylov subspaces are chosen in such a way that they are not part of the previous Krylov subspaces. This means

  $$v_j \notin \bigcup_{i=1}^{j-1} \mathscr{K}_{l_i}(A, v_i).$$

In the following two subsections we will note what changes in the theoretical derivations for obtaining the necessary and sufficient conditions are required. In a final subsection we will investigate in more detail the case of deflation.

5.2. The necessary conditions

For the derivation of the necessary conditions we assumed, for simplicity reasons that we were working with only one Krylov subspace. In the case of invariant subspaces however, we have to work with a union of these Krylov subspaces. This means that our orthogonal transformations $Q_{0:k-1}$ satisfy the following conditions:

(1) The columns of $\overleftarrow{Q}_{0:k-1}$ form an orthogonal basis for
$K_k(A) = \mathscr{K}_{l_1}(A, v_1) \cup \cdots \cup \mathscr{K}_{k_t}(A, v_t)$.

(2) The columns of $\overrightarrow{Q}_{0:k-1}$ form an orthogonal basis for the orthogonal complement of $K_k(A)$.

In fact this is the only thing which changes in these derivations, all the remaining statements stay valid. The conditions put on the orthogonal matrices Q_k and on the matrices $\widetilde{A}^{(k)}$ remain valid.

However in the case of invariant subspaces we can derive also the following property. The occurrence of invariant subspaces, creates zero blocks below the diagonal in the matrices $A^{(k)}$. Suppose that for a certain k the space $K_k(A)$ becomes invariant. Due to the invariance we have: $AK_k(A) \subset K_k(A)$. As the matrix $\overleftarrow{Q}_{0:k-1}$ forms an orthogonal basis for the space $K_k(A)$, we get the following equations:

$$\begin{aligned} Q_{0:k-1}^{\mathrm{T}} A Q_{0:k-1} &= A^{(k)}, \\ A Q_{0:k-1} &= Q_{0:k-1} A^{(k)} \\ &= Q_{0:k-1} \left(\begin{array}{c|c} R_k & \times \\ \hline 0 & A_k \end{array} \right). \end{aligned}$$

In case of an invariant subspace the matrix $A^{(k)}$ has a zero block of dimension $(n-k) \times k$ in the lower left position. In fact one can apply deflation now and continue working with the lower right block A_k, e.g. for finding eigenvalues. If one applies deflation after every invariant subspace, one does in fact not work with the whole space $K_k(A)$, but on separate (invariant) subspaces. Indeed, after the first invariant subspace $K_{l_1}(A) = \mathscr{K}_{l_1}(A, v_1)$ one applies deflation and one starts iterating this procedure on a new matrix A_{l_1}.

5.3. Sufficient conditions

In this subsection, we will take a closer look at the proof of Theorem 2, and investigate the changes in case of an invariant Krylov subspace.

Suppose at step k of the reduction algorithm we encounter an invariant subspace $K_k(A)$, i.e. $K_k(A) = \mathscr{K}_{l_1}(A, v_1) \cup \cdots \cup \mathscr{K}_{l_t}(A, v_t)$. Only the last lines of the proof of Theorem 2, do not hold anymore. The following equation however remains valid:

$$K_k(A) \subset \mathrm{Range}(\overleftarrow{Q}_{0:k}).$$

This leads to the following equation:

$$\left(\overleftarrow{Q}_{0:k} | \overrightarrow{Q}_{0:k} \right) = \left(\overleftarrow{Q}_{0:k-1} | v_{t+1} | \overrightarrow{Q}_{0:k} \right) \left(\begin{array}{c|c} W & 0 \\ \hline 0 & I \end{array} \right)$$

with W a matrix of dimension $(k+1) \times (k+1)$. Therefore

$$\mathrm{Range}(\overleftarrow{Q}_{0:k}) = K_k(A) \cup \langle v_{t+1} \rangle$$

with

$$v_{t+1} \notin \text{Range}(\overleftarrow{Q}_{0:k-1}) = \bigcup_{i=1}^{t} \mathscr{K}_{l_i}(A, v_i).$$

So, defining $K_{k+1}(A) = \mathscr{K}_{l_1}(A, v_1) \cup \cdots \cup \mathscr{K}_{l_t}(A, v_t) \cup \mathscr{K}_1(A, v_{t+1})$ proves the theorem.

One might wonder if it is possible to choose any vector v not in $K_k(A)$, for defining the new space $K_{k+1}(A)$, because our matrix Q_k^{T} still has to satisfy some conditions. Let us define the vector w as the orthogonal projection of the vector v onto $\overrightarrow{Q}_{0:k-1}$. Then we have that

$$K_k(A) \cup \langle v \rangle = K_k(A) \cup \langle w \rangle \tag{3}$$

with w orthogonal to $\overleftarrow{Q}_{0:k-1}$. We have the relation:

$$(\overleftarrow{Q}_{0:k} | \overrightarrow{Q}_{0:k}) Q_k^{\mathrm{T}} = (\overleftarrow{Q}_{0:k-1} | \overrightarrow{Q}_{0:k-1}),$$

so if we choose now $\overleftarrow{Q}_{0:k} = (\overleftarrow{Q}_{0:k-1} | w)$, we see that the conditions put on Q_k and $\widetilde{A}^{(k)}$ still are satisfied and moreover we have chosen v to be an arbitrary vector not in $K_k(A)$, because the matrix $\overleftarrow{Q}_{0:k}$ spans the space $K_{k+1}(A) = K_k(A) \cup \langle v \rangle$, due to Eq. (3).

5.4. *The case of deflation*

Finally we will take a closer look at the case of deflation. Suppose now that for a certain k we get an invariant space $K_k(A) = \mathscr{K}_{l_1}(A, v_1) \cup \cdots \cup \mathscr{K}_{l_t}(A, v_t)$. Suppose we apply deflation and we can continue iterating the procedure on the matrix A_k. Assume we start on this matrix the procedure with unit starting vector w, by applying an initial transformation W_2. This means that $W_2 e_1 = w$. In fact this corresponds to applying the similarity transformation (with W_1 an arbitrary orthogonal matrix):

$$Q_k = \left(\begin{array}{c|c} W_1 & 0 \\ \hline 0 & W_2 \end{array}\right),$$

which satisfies the desired condition (the upper right $k \times (n-k-1)$ block is zero), on the matrix $A^{(k)}$. Looking closer at the matrix $Q_{0:k}$ we get the following relations:

$$\begin{aligned} Q_{0:k} &= Q_{0:k-1} Q_k \\ &= (\overleftarrow{Q}_{0:k-1} | \overrightarrow{Q}_{0:k-1}) Q_k. \end{aligned}$$

This means that

$$\overleftarrow{Q}_{0:k} = (\overleftarrow{Q}_{0:k-1} W_1 | v_{t+1}).$$

Clearly the vector $v_{t+1} = \overrightarrow{Q}_{0:k-1} w$ does not belong to the space $K_k(A)$, moreover the vector is perpendicular to $K_k(A)$. This last deduction clearly shows the relation between the different Krylov spaces when applying deflation and the continuation of the reduction process on the complete matrix.

6. Some general remarks

When we take a closer look at the matrix equation:

$$Q_k^{\mathrm{T}} = Q_{0:k}^{\mathrm{T}} Q_{0:k-1}$$

$$= \begin{pmatrix} \overleftarrow{Q}_{0:k}^{\mathrm{T}} \\ \overrightarrow{Q}_{0:k}^{\mathrm{T}} \end{pmatrix} \left(\overleftarrow{Q}_{0:k-1} | \overrightarrow{Q}_{0:k-1} \right),$$

we can see that the matrix Q_k^{T} has the upper right $(k+1) \times (n-k)$ block of rank less than or equal to 1. The upper right $(k+1) \times (n-k)$ block corresponds to the product $\overleftarrow{Q}_{0:k}^{\mathrm{T}} \overrightarrow{Q}_{0:k-1}$. The columns of the matrix $\overleftarrow{Q}_{0:k}^{\mathrm{T}}$ span the subspace $K_{k+1}(A) = K_k(A) + \langle A^{k_t} v_t \rangle$ (assuming that $K_k(A)$ is not invariant) and the columns of $\overrightarrow{Q}_{0:k-1}$ span the space orthogonal to $K_k(A)$, which leads directly to the fact that the product $\overleftarrow{Q}_{0:k}^{\mathrm{T}} \overrightarrow{Q}_{0:k-1}$, has rank less than or equal to 1. The invariant case can be dealt with in a similar way.

The reader can easily verify that the similarity reductions of a symmetric matrix into a similar tridiagonal or a semiseparable one, and the similarity reduction of a matrix into a similar Hessenberg or a matrix having the lower triangular part of semiseparable form [9], perfectly fit in this scheme. Moreover one can derive that the vector v equals e_1, if of course the initial transformation Q_0 equals the identity matrix.

7. Conclusions

In this paper we derived two easy conditions satisfied by orthogonal similarity transformations, such that the resulting partially reduced matrices have in the already reduced part the Lanczos (Arnoldi)–Ritz values as eigenvalues. Moreover we proved that these conditions are necessary and sufficient.

References

[1] J.K. Cullum, R.A. Willoughby, Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Birkhäuser, Boston, 1985.

[2] J.W. Demmel, Applied Numerical Linear Algebra, SIAM, 1997.

[3] D. Fasino, L. Gemignani, A Lanczos type algorithm for the QR-factorization of regular Cauchy matrices, Numer. Linear Algebra Appl. 9 (2002) 305–319.

[4] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., The Johns Hopkins University Press, 1996.

[5] A.B.J. Kuijlaars, Which eigenvalues are found by the Lanczos method? SIAM J. Matrix Anal. Appl. 22 (1) (2000) 306–321.

[6] B.N. Parlett, The symmetric eigenvalue problem, Classics in Applied Mathematics, SIAM, Philadelphia, 1998.

[7] Y. Saad, Numerical Methods for Large Eigenvalue Problems, Manchester University Press, Manchester, UK, 1992.

[8] L.N. Trefethen, D. Bau, Numerical Linear Algebra, SIAM, 1997.

[9] M. Van Barel, R. Vandebril, N. Mastronardi, An orthogonal similarity reduction of a matrix into semiseparable form, SIAM J. Matrix Anal. Appl. 27 (1) (2005) 176–197.

[10] R. Vandebril, Semiseparable matrices and the symmetric eigenvalue problem, PhD thesis, Dept. of Computer Science, K.U. Leuven, Celestijnenlaan 200A, 3000 Leuven, May 2004.

[11] R. Vandebril, E. Van Camp, M. Van Barel, N. Mastronardi, An orthogonal similarity reduction of a matrix into semiseparable plus diagonal form, with a free choice of the diagonal, Technical Report TW398, Katholieke Universiteit Leuven, Dept. Computerwetenschappen, Celestijnenlaan 200A, 3001 Heverlee (Leuven), Numerische Mathematik, submitted for publication.