Necessary and sufficient conditions for orthogonal similarity transformations to obtain the Ritz values

Raf Vandebril
Marc Van Barel

Report TW 396, July 2004

Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

Necessary and sufficient conditions for orthogonal similarity transformations to obtain the Ritz values

Raf Vandebril
Marc Van Barel

Report TW 396, July 2004

Department of Computer Science, K.U.Leuven

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 (see e.g. [Golub G. and Van Loan C.]). This behavior is also shared by the reduction algorithm which transforms symmetric matrices via orthogonal similarity transformations to semiseparable form (see [Van Barel, Vandebril, Mastronardi]). 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 which have to be placed on the orthogonal similarity transformation, 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 : inverse of Hessenberg matrices, semiseparable matrices, implicit Q-theorem, Hessenberg-like matrices.
AMS(MOS) Classification : Primary : 65F15, Secondary : 15A18.

Necessary and sufficient conditions for orthogonal similarity transformations to obtain the Ritz values *

Raf Vandebril †, Marc Van Barel ‡

13th July 2004

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 (see e.g. [4]). This behavior is also shared by the reduction algorithm which transforms symmetric matrices via orthogonal similarity transformations to semiseparable form (see [9]). 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 which have to be placed on the orthogonal similarity transformation, 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.

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, 7, 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 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 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 proving 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 paper is organized as follows. In Section 2 we introduce briefly the type of similarity transformation considered, and also the notion of Ritz values and Krylov subspaces is briefly refreshed. The similarity transformations obeying the desired convergence behavior are investigated in Section 3. This leads to two

*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 & Modelling). The scientific responsibility rests with the authors.

†Email:raf.vandebril@cs.kuleuven.ac.be

‡Email:marc.vanbarel@cs.kuleuven.ac.be

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. Some general remarks, and an extra property of the orthogonal matrices are derived in Section 5. 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 $A^{(1)} = Q_0^T A^{(0)} Q_0$. We remark that the transformation Q_0 is not essential. Quite often the orthogonal matrix Q_0 is equal to the identity matrix. E.g., in the reduction to semiseparable or tridiagonal form the matrix Q_0 equals the identity matrix I. It is however perfectly possible to perform a first transformation Q_0 on the matrix $A^{(0)}$. This will not affect the reduction algorithms, but it will affect the convergence behavior of the reduction, as we will show in this subsection. The other orthogonal transformations Q_k, $k \geq 1$, 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

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

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:

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

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 dimension k with initial vector v:

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

where $\langle x, y, z \rangle$ denotes the vector space spanned by the vectors x, y and z. If the columns of the matrix $\overleftarrow{Q}_{0:k}$ form an orthonormal basis of the Krylov subspace $\mathcal{K}_{k+1}(A, v)$, 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 $A^{(1)} = Q_0^T A Q_0$ to the desired form.

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

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

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

After the next step in the transformation we have that the block R_{k+1} has the Arnoldi-Ritz values as eigenvalues with respect to $\mathcal{K}_{k+1}(A, v)$. 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 $\mathcal{K}_{k+1}(A, v) = \mathcal{K}_k(A, v) + \langle A^k v \rangle$.
2. The columns of $\overrightarrow{Q}_{0:k}$ form an orthogonal basis for the orthogonal complement of $\mathcal{K}_{k+1}(A, v)$.

We have the following equalities:

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

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

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

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

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

where the $(Q_k)_{11}^T, (Q_k)_{12}^T, (Q_k)_{21}^T$ and $(Q_k)_{22}^T$ denote a partitioning of the matrix Q_k^T. These blocks have the following dimensions: $(Q_k)_{11}^T \in \mathbb{R}^{(k+1)\times k}$, $(Q_k)_{12}^T \in \mathbb{R}^{(k+1)\times(n-k)}$, $(Q_k)_{21}^T \in \mathbb{R}^{(n-k-1)\times k}$ and $(Q_k)_{22}^T \in \mathbb{R}^{(n-k-1)\times(n-k)}$. 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 $(Q_k)_{21}^T$ 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 matrix $\tilde{A}^{(k)}$ satisfying

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

which can be rewritten as:

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

Rewriting equation (1) gives us:

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

Because the columns of $A\overleftarrow{Q}_{0:k-1}$ belong to the Krylov subspace: $\mathcal{K}_{k+1}(A,v)$, which is spanned by the columns of $\overleftarrow{Q}_{0:k}$, we have that $\tilde{A}^{(k)}$ 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 $\tilde{A}^{(k)}$ 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:*

- *The matrix Q_k^T, which is the orthogonal matrix to transform $A^{(k)}$ into the matrix $A^{(k+1)} = Q_k^T A^{(k)} Q_k$ has a zero block of dimension $(n-k-1)\times k$ in the lower left corner.*
- *The matrix $\tilde{A}^{(k)} = Q_k^T A^{(k)}$ 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 $\tilde{A}^{(k)}$ are sufficient to have the Arnoldi-Ritz values as eigenvalues in the blocks A_k.

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

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

Suppose that at step k 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 $A^{(k+1)} = Q_k^T A^{(k)} Q_k$ has a zero block of dimension $(n-k-1)\times k$ in the lower left corner;*
- *the matrix $\tilde{A}^{(k)} = Q_k^T A^{(k)}$ has a zero block of dimension $(n-k-1)\times k$ in the lower left corner.*

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

$$A^{(k)} = \left(\begin{array}{c|c} R_k & \times \\ \hline \times & A_k \end{array}\right),$$

that the matrix R_k of dimension $k\times k$ has the Arnoldi-Ritz values with respect to the Krylov space $\mathcal{K}_k(A,v)$.

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

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

Step k Suppose the theorem is true for $A^{(1)}, A^{(2)}, \ldots, A^{(k-1)}$. This means that the columns of $\overleftarrow{Q}_{0:k-1}$ span the Krylov subspace $\mathcal{K}_k(A, v)$. 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 $\mathcal{K}_{k+1}(A, v)$. We have the following equalities

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

Therefore,

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

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 $A\mathcal{K}_k(A, v)$. We have the following relation:

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

With Range(A) we denote the vector space spanned by the columns of the matrix A. Since Q_k^T has a zero block in the lower left position, we have that:

$$\begin{aligned} Q_{0:k} &= Q_{0:k-1} Q_k \\ Q_{0:k} Q_k^T &= Q_{0:k-1}.\end{aligned}$$

Hence,

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

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

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

When we combine this, with equation (2) we get:

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

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

□

5 Some general remarks

When we take a closer look at the matrix equation:

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

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 $\overleftarrow{Q}_{0:k}^T \overrightarrow{Q}_{0:k-1}$. The columns of the matrix $\overleftarrow{Q}_{0:k}^T$ span the Krylov subspace $\mathcal{K}_{k+1}(A, v) = \mathcal{K}_k(A, v) + \langle A^k v \rangle$ and the columns of $\overrightarrow{Q}_{0:k-1}$ span the space orthogonal to $\mathcal{K}_k(A,v)$, which leads directly to the fact that the product $\overleftarrow{Q}_{0:k}^T \overrightarrow{Q}_{0:k-1}$, has rank less than or equal to 1. 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.

6 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 and 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 and L. Gemignani. A Lanczos type algorithm for the *QR*-factorization of regular Cauchy matrices. *Numerical Linear Algebra with Applications*, 9:305–319, 2002.

[4] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins university Press, third edition, 1996.

[5] A.B.J. Kuijlaars. Which eigenvalues are found by the Lanczos method? *SIAM Journal on Matrix Analysis and its Applications*, 22(1):306–321, 2000.

[6] B.N. Parlett. *The symmetric eigenvalue problem*, volume 20 of *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 and D. Bau. *Numerical Linear Algebra*. SIAM, 1997.

[9] M. Van Barel, R. Vandebril, and N. Mastronardi. The Lanczos-Ritz values appearing in an orthogonal similarity reduction of a matrix into semiseparable form. Report TW 360, Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, B-3001 Leuven (Heverlee), May 2003.