A Wilkinson-like multishift QR algorithm for symmetric eigenvalue problems and its global convergence

Kensuke Aishima*, Takayasu Matsuo, Kazuo Murota, Masaaki Sugihara

University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, Japan

ARTICLE INFO

Article history:
Received 27 September 2010
Received in revised form 25 October 2010

MSC:
65F15

Keywords:
Numerical linear algebra
Eigenvalue
Symmetric tridiagonal matrix

ABSTRACT

In 1989, Bai and Demmel proposed the multishift QR algorithm for eigenvalue problems. Although the global convergence property of the algorithm (i.e., the convergence from any initial matrix) still remains an open question for general nonsymmetric matrices, in 1992 Jiang focused on symmetric tridiagonal case and gave a global convergence proof for the generalized Rayleigh quotient shifts. In this paper, we propose Wilkinson-like shifts, which reduce to the standard Wilkinson shift in the single shift case, and show a global convergence theorem.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Modern eigenvalue computation methods usually consist of the following two phases: a given matrix is first transformed to a Hessenberg matrix, and then its eigenvalues are computed by some iterative method. As the iterative method, the QR algorithm [1–4] had been the most important tool for many years in single processor environments. In 1989, Bai–Demmel extended the QR algorithm to the multishift QR algorithm [5], which exploits multiple shifts at the same time on different processors to enhance the performance (in view of this, the original QR algorithm is now called the “single shift” QR). It was then followed by the excellent implementation in [6,7], which is now widely used as the DHSEQR routine in LAPACK. Furthermore, some efficient multishift implementations for several special matrices have been developed [8–10].

In spite of the success, one possible drawback of the QR algorithms is that theoretical aspects of the convergence have not yet been fully investigated, except for a few local convergence results, i.e., convergence in a certain single step; for example, in the case of the generalized Rayleigh quotient shifts, which are the most typical choices in the multishift QR algorithm, it has been shown that the error typically decreases quadratically [11]. We like to emphasize here that there is no theorem on the global convergence property (i.e., convergence from any given matrix). It seems that the set of nonsymmetric matrices is still too large to establish such theorems, as far as the authors understand, and further basic studies are demanded.

The present paper belongs to such basic theoretical studies. In this paper we limit ourselves to the symmetric case, for which several global convergence results are known, and we fill a missing part in the existing studies.

Before stating our point more specifically, we define our notation and briefly summarize the existing theories. The elements of an $m \times m$ irreducible symmetric tridiagonal matrix $T$ is defined by

$$
T = \begin{pmatrix}
\alpha_1 & \beta_1 \\
\beta_1 & \alpha_2 & & \\
& \ddots & \ddots & \\
& & \beta_{m-1} & \alpha_m
\end{pmatrix},
$$

(1)

* Corresponding author.
E-mail address: Kensuke_Aishima@mist.i.u-tokyo.ac.jp (K. Aishima).

1 See [12–14] for the symmetric version of the multishift QR algorithm.

0377-0427/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
The multishift QR algorithm is described as
\[
(T_n - s_1^{(n)} I) \cdots (T_n - s_l^{(n)} I) = Q_n R_n,
\]
where \(Q_n\) is an orthogonal matrix, \(R_n\) is an upper triangle matrix, and \(T_0 := T\). The shifts \(s_1^{(n)}, \ldots, s_l^{(n)}\) are assumed to be taken by some given procedure. Similarly for (1), let us write the matrix \(T_n\) as
\[
T_n = \begin{pmatrix}
\alpha_1^{(n)} & \beta_1^{(n)} \\
\beta_1^{(n)} & \alpha_2^{(n)} & \beta_2^{(n)} \\
& \ddots & \ddots \\
& & \ddots & \beta_{m-1}^{(n)} \\
& & & \alpha_m^{(n)}
\end{pmatrix}.
\]

There is a long history of convergence analysis for tridiagonal QR algorithms [15–20]. The important results on the global convergence are summarized in Table 1.

First, we summarize the results on the single shift QR algorithm (\(l = 1\); see also [19] for a review). In case of the single Rayleigh quotient shift (\(s^{(n)} = \alpha_m^{(n)}\), the global convergence theorem reads as follows.

**Theorem 1** (Wilkinson [20]). Suppose the QR algorithm with the Rayleigh quotient shift is applied to \(T\). Then we have the following facts:

(a) \(\{|\beta_{m-1}^{(n+1)}|\}\) is a monotone nonincreasing sequence.
(b) \(\lim_{n \to \infty} |\beta_{m-1}^{(n)}| = 0\) and/or \(\lim_{n \to \infty} |\beta_{m-2}^{(n)}| = 0\).

Theorem 1 states that either or both of the lower-right elements \(\beta_{m-1}^{(n)}\) and \(\beta_{m-2}^{(n)}\) tend(s) to 0 for any given \(T\), which means a global convergence in view of deflation. However, in the same paper Wilkinson [20] also showed that in some initial matrices, the convergence rate is only linear. This gave him a motivation to make a better shift, the Wilkinson shift, where the lower right 2 × 2 submatrix is picked and its eigenvalue closer to \(\alpha_m^{(n)}\) is used as a shift. The associated global convergence theorem reads as follows.

**Theorem 2** (Wilkinson [20]). Suppose the QR algorithm with the Wilkinson shift is applied to \(T\). Then we have the following facts:

(a) \(\{|\beta_{m-2}^{(n+1)} | \beta_{m-1}^{(n+1)}|\}\) is a monotone nonincreasing sequence.
(b) \(\lim_{n \to \infty} |\beta_{m-2}^{(n)}| = 0\).
(c) For any \(\varepsilon > 0\), there exists an \(n\) such that \(|\beta_{m-1}^{(n)}| < \varepsilon\).

The shift is better in that it always achieves \(\lim_{n \to \infty} |\beta_{m-1}^{(n)}| = 0\), and furthermore the rate is quadratic (see [20] for details; see also [16] for another proof of \(\lim_{n \to \infty} |\beta_{m-1}^{(n)}| = 0\)).

Next let us consider the multishift QR algorithm. Usually, the shifts \(s_1^{(n)}, \ldots, s_l^{(n)}\) are taken to be the eigenvalues of the lower right \(l \times l\) matrix of \(T^{(n)}\). This is the so-called generalized Rayleigh quotient shift strategy. In 1992, Jiang gave a convergence theorem for this algorithm, as a natural multishift extension of Theorem 1.

**Theorem 3** (Jiang [17]). Let \(l \leq (m - 1)/2\). Suppose the multishift QR algorithm is applied to \(T\) and \(s_1, \ldots, s_l\) are the generalized Rayleigh quotient shifts. Then we have the following facts:

(a) \(\{|\beta_{m-1}^{(n+1)} \cdots \beta_{m-1}^{(n+1)}|\}\) is a monotone nonincreasing sequence.
(b) \(\lim_{n \to \infty} |\beta_{m-1}^{(n)} \cdots \beta_{m-1}^{(n)}| = 0\) and/or \(\lim_{n \to \infty} |\beta_{m-2}^{(n)} \cdots \beta_{m-2}^{(n)}| = 0\).

From Theorem 3, we see
\[
\lim_{n \to \infty} \min_{1 \leq i \leq 2l} \beta_{m-i}^{(n)} = 0,
\]
which implies the global convergence of the algorithm.
At this point, we naturally think that in a similar manner a multishift version of the Wilkinson shift can be found by simply taking a larger submatrix; in fact, the idea was mentioned in [21]. So far, however, no concrete strategy has been proposed, to the best of the authors’ knowledge. In this respect, the lower right part of Table 1 means not only the lack of a global convergence theorem, but more essentially the absence of such a shift strategy itself. This might seem surprising, but one reason for this is that although the proposal of such a strategy is straightforward, we need further consideration and modification of the algorithm in order to establish a global convergence result.

In view of the background, in the present paper we propose a Wilkinson-like multishift QR algorithm, and prove its global convergence. The point here is to carefully choose the size of the submatrix so that the global convergence result follows. Theorem 6 is the main result, which corresponds to Theorem 2 (single shift case). This completes Table 1. We like to emphasize here that this paper is mainly of theoretical interest, as Jiang [17].

Throughout this paper let \( q_k^{(n)} \) be the \( k \)th vector of \( Q_n \) in (2) and \( q_k^{(m)} \) be the \( (k_1, k_2) \) element of \( Q_m \). Furthermore, let \( E_m \) be the \( m \times m \) identity matrix and \( e_k \) be the \( k \)th vector of \( E_m \).

2. A Wilkinson-like multishift

In this section we propose the Wilkinson-like multishift. Let \( l \) denote the number of shifts and let \( \lambda_1^{(n)} > \cdots > \lambda_{l+1}^{(n)} \) denote the eigenvalues of the lower right \((l+1) \times (l+1)\) submatrix of \( T_n \). In the present multishift strategy, those eigenvalues are computed, and then the eigenvalues except the one satisfying

\[
\frac{|\lambda_{m-1}^{(n)}|}{|\alpha_m^{(n)} - \lambda_k^{(n)}|} \leq 1
\]

(6)

are chosen as shifts. Since this shift strategy reduces to the standard Wilkinson shift strategy when \( l = 1 \), we call it a Wilkinson-like shift strategy for the multishift QR algorithm in this paper.

Remark 4. The reader might wonder whether or not we can actually find an eigenvalue \( \lambda_k^{(n)} \) satisfying (6) for every step \( n \), but this can be easily accomplished in the following way. At step \( n \), let us consider the eigenvalue \( \lambda_k^{(n)} \) that maximizes \( |\lambda_{k'} - \alpha_{k'}^{(n)}| (k' = 1, \ldots, l+1) \); then this automatically satisfies (6) (although this might not be the only choice). In order to see this, let us consider the eigenvalues of the lower right \( 2 \times 2 \) submatrix, and let \( \mu^{(n)} \) be the eigenvalue which is farther away from the lower right element \( \alpha_{m}^{(n)} \). Then \( \mu^{(n)} \) satisfies \( |\lambda_{m-1}^{(n)}/(\alpha_{m}^{(n)} - \mu^{(n)})| \leq 1 \) [20]. In view of \( \lambda_{m}^{(n)} \geq \mu^{(n)} \geq \lambda_{k+1}^{(n)} \), (6) is satisfied. \( \square \)

3. Convergence theorem

In this section we prove global convergence of the Wilkinson-like multishift QR algorithm. We first prepare the following lemma for tridiagonal matrices.

Lemma 5 ([19, Theorem 7.3.1]). For a symmetric tridiagonal matrix \( T \) whose elements are defined as (1), let \( \lambda_1, \ldots, \lambda_l \) be the eigenvalues of the lower right \( l \times l \) submatrix, \( e_{m-1} \) be the \( (m-l) \)th column of the identity matrix \( E_m \). Then the \( m \)th column of the matrix product \( (T - \lambda_1 I) \cdots (T - \lambda_l I) = \beta_{m-l} \cdots \beta_{m-1} e_{m-1} \). Similarly, the \( m \)th row is \( \beta_{m-1} \cdots \beta_{m-l} e_{m-1}^T \).

Based on the lemma above, a global convergence theorem of the Wilkinson-like multishift QR algorithm is obtained.

Theorem 6. Let \( l \leq m - 2 \). Suppose the multishift QR algorithm is applied to \( T \) and \( s_1, \ldots, s_l \) are the Wilkinson-like multishift. Then we have the following facts:

(a) \(|\beta_{m-1}^{(n)} \cdots \beta_{m-l}^{(n)}|\) is a monotone nonincreasing sequence.

(b) \( \lim_{n \to \infty} |\beta_{m-1}^{(n)} \cdots \beta_{m-l}^{(n)}| = 0 \).

(c) For any \( \epsilon > 0 \), there exists an \( n \) such that \(|\beta_{m-1}^{(n)} \cdots \beta_{m-l}^{(n)}| < \epsilon \).

Proof. Our proof is similar to [17]. From (2), we see

\[
(T_n - s_{k_1}^{(n)}) \cdots (T_n - s_{k_-1}^{(n)}) Q_n = R_n^T.
\]

(7)

It then follows that

\[
(T_n - \lambda_1^{(n)} I) \cdots (T_n - \lambda_l^{(n)} I) Q_n = (T_n - \lambda_k^{(n)} I) R_n^T
\]

(8)

is satisfied for the Wilkinson-like multishift QR algorithm. In view of the \( m \)th column of the \( m \times m \) matrix in the above equality, we have

\[
(T_n - \lambda_{l+1}^{(n)} I) \cdots (T_n - \lambda_1^{(n)} I) e_m = (T_n - \lambda_1^{(n)} I) r_m^{(n)} e_m.
\]

(9)
where $q_{m}^{(n)}$ is the $m$th column of $Q$, $e_{m}$ is the $m$th column of $E_{m}$, and $r_{mm}(n)$ is the $(m, m)$ element of $R_{n}$. Moreover, the $m$th element of the equality above and Lemma 5 imply

$$
\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)} q_{m-l+1,m}^{(n)} = (e_{m}^{(n)} - \lambda_{k}^{(n)}) r_{mm}^{(n)}.
$$

(10)

Since the $(m - l + 1, m)$ element of $T_{n+1}^{(n)}$ is $\rho_{m-l+1}^{(n)} \cdots \rho_{m-1}^{(n)}$, we see

$$
\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)} = e_{m-l+1}^{T} (T_{n+1} - s_{1}^{(n)}) \cdots (T_{n+1} - s_{l}^{(n)}) e_{m},
$$

and

$$
= q_{m-l+1}^{(n)} T_{n+1} - s_{1}^{(n)} \cdots (T_{n+1} - s_{l}^{(n)}) q_{m-l+1}^{(n)} e_{m},
$$

$$
= q_{m-l+1}^{(n)} T_{n+1} - s_{l}^{(n)} q_{m}^{(n)},
$$

$$
= q_{m-l+1}^{(n)} T_{n+1} - s_{l}^{(n)} q_{m-l+1}^{(n)} e_{m}.
$$

we see

$$
\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)} = \beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)} T_{n+1} - s_{l}^{(n)} q_{m-l+1}^{(n)} e_{m}.
$$

(11)

where the first equality is due to $e_{m-l+1}^{T} T_{n+1}^{(n)} e_{m} = 0$ ($l' < l + 1$), the second equality is due to (3), the third equality is due to (9), the fourth equality is due to $q_{m-l+1}^{(n)} = 0$, and the last equality is due to (10).

Therefore, we have

$$
|\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)}| \leq |\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)}|
$$

from (6).

Next we prove (b). From the equality above, $|\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)}|$ is convergent. Let $\lim_{n \to \infty} |\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)}| = C$. Actually, $C = 0$ is proved by contradiction as follows. If $C > 0$, then

$$
\lim_{n \to \infty} q_{m-l+1,m}^{(n)} = 1, \quad \lim_{n \to \infty} q_{m-l-1,m}^{(n)} = 1, \quad \lim_{n \to \infty} \beta_{m-1}^{(n)} = 1
$$

from (11). Hence,

$$
\lim_{n \to \infty} q_{m-l, m}^{(n)} = 0 \quad (l' = 0, 1, \ldots, l).
$$

From the $m$th column of the $m \times m$ matrix in the equality (7), we see

$$
(T_{n} - s_{1}^{(n)} I) \cdots (T_{n} - s_{l}^{(n)} I) q_{m}^{(n)} = r_{mm}^{(n)} e_{m}.
$$

(14)

The $(m, m - l')$ elements of $(T_{n} - s_{1}^{(n)} I) \cdots (T_{n} - s_{l}^{(n)} I)$ are 0 with $l' > l$, so $\lim_{n \to \infty} r_{mm}^{(n)} = 0$. Therefore, we have

$$
\lim_{n \to \infty} |\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)}| = 0
$$

from (10). This contradicts $C > 0$. Thus we have

$$
\lim_{n \to \infty} |\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)}| = 0.
$$

(15)

Finally we show (c). Since we see

$$
\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)} = e_{m-l}^{T} T_{n+1} e_{m},
$$

similarly to (11), we have

$$
\beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)} = e_{m-l}^{T} (T_{n+1} - s_{1}^{(n)} I) \cdots (T_{n+1} - s_{l}^{(n)} I) e_{m}
$$

$$
= q_{m-l+1}^{(n)} T_{n+1} - s_{1}^{(n)} \cdots (T_{n+1} - s_{l}^{(n)} I) q_{m-l+1}^{(n)} e_{m},
$$

$$
= q_{m-l+1}^{(n)} T_{n+1} - s_{l}^{(n)} q_{m-l+1}^{(n)} e_{m},
$$

$$
= q_{m-l+1}^{(n)} T_{n+1} - s_{l}^{(n)} q_{m-l+1}^{(n)} e_{m},
$$

$$
= q_{m, m-l}^{(n)} e_{m-l+1,m}^{(n)} \beta_{m-1}^{(n)} \cdots \beta_{m-l+1}^{(n)}.
$$

(16)
where the first equality is due to $\mathbf{e}_{m-1}^T T_{n+1} \mathbf{e}_m = 0$ ($l < l'$), the second equality is due to (3), the third equality is due to (14), and the last equality is due to (10). It then follows that
\[
|\beta_{m-1}^{(n+1)} \cdots \beta_{m-1}^{(n+1)}| \leq |\beta_{m-2}^{(n)} \cdots \beta_{m-1}^{(n-1)}|
\]
from (6). In this inequality $|\beta_{m-2}^{(n)} \cdots \beta_{m-1}^{(n)}|$ is bounded: $|\beta_{m-2}^{(n)} \cdots \beta_{m-1}^{(n-1)}| \leq M$ (note that $T_n$ has the same eigenvalues as $T$, and hence they are bounded; this implies all elements of $T_n$ are also bounded). Hence, we see
\[
|\beta_{m-1}^{(n)} \cdots \beta_{m-1}^{(n)}| \leq |\beta_{m-2}^{(n)} \cdots \beta_{m-1}^{(n-1)}| \leq M|\beta_{m-1}^{(n)} \cdots \beta_{m-1}^{(n-1)}|.
\]
From (15), for any $\epsilon > 0$, $|\beta_{m-1}^{(n)} \cdots \beta_{m-1}^{(n)}| < \epsilon$ or $|\beta_{m-2}^{(n)} \cdots \beta_{m-1}^{(n-1)}| < \epsilon$ for all sufficiently large $n$. From (16) we have
\[
\lim_{n \to \infty} |\beta_{m-1}^{(n)} \cdots \beta_{m-1}^{(n)}| = 0.
\]
Therefore, (c) is obtained. \(\square\)

Notice that the property (17) in the proof is stronger than the statement (c) in Theorem 6 (which was stated in that form so that the relation with Theorem 2 by Wilkinson can be easily seen). A stronger result reflecting this property can be written as follows.

**Corollary 7.** Suppose $s_1, \ldots, s_l$ are the Wilkinson-like multishifts. Then we have
\[
\lim_{n \to \infty} \min_{1 \leq i' \leq l} |\beta_{m-l'}^{(n)}| = 0,
\]
\[
\lim_{n \to \infty} \min_{1 \leq i' \leq l} |\beta_{m-l'}^{(n)}| = 0. \quad \square
\]

This implies that for any given matrix the iteration of the Wilkinson-like multishift QR algorithm always eventually comes to a deflation. This is the desired global convergence result.

**Acknowledgments**

The authors thank the anonymous referees for valuable comments which helped improving the quality of this paper. The authors are thankful to Prof. Van Barel for helpful comments in ICCAM-2010. This study is partially supported by the Global Center of Excellence “the research and training center for new development in mathematics”. The first author is supported by the Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists.

**References**


