MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

A Complete Analysis of Convergence Rate of the Tridiagonal QR Algorithm with Wilkinson's Shift

Kensuke AISHIMA, Takayasu MATSUO,
Masaaki SUGIHARA

METR 2013–26 October 2013

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

A Complete Analysis of Convergence Rate of the Tridiagonal QR Algorithm with Wilkinson's Shift

Kensuke AISHIMA[1], Takayasu MATSUO[1], Masaaki SUGIHARA[2]

[1]Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo
{Kensuke_Aishima,matsuo}@mist.i.u-tokyo.ac.jp

[2]Department of Phisics and Mathematics,
Aoyama Gakuin University,
sugihara@gem.aoyama.ac.jp

October 2013

Abstract

We discuss the convergence rate of the QR algorithm with Wilkinson's shift for tridiagonal symmetric eigenvalue problems. It is well known that the convergence rate is theoretically at least quadratic, and practically cubic for most matrices. In an effort to separate the quadratic/cubic cases, a standard classification method based on the limiting matrices has been established. In this paper, we show by an example that there still remains a convergence scenario not mentioned in this classification, and give a new "complete" classification covering all the possible scenarios.

1 Introduction

The standard method for computing eigenvalues of a symmetric matrix A has two steps. First, A is transformed to a tridiagonal matrix T by an appropriate orthogonal transformation. Then some iterative method is applied to T to compute its eigenvalues. There are several approaches in the second step. Among them, the historical QR algorithm is still widely used as a reliable tool, particularly with Wilkinson's shift for accelerating the convergence. In this paper we consider the convergence behavior of this algorithm.

For this algorithm, there is a long history of convergence analysis. Global convergence was first proved by Wilkinson in 1968 [7], and then another elegant proof was given by Hoffmann–Parlett [1]. Regarding the convergence rate, Wilkinson [7] theoretically proved that it is at least quadratic, after which Hoffmann–Parlett [1] showed that in most matrices better cubic convergence is achieved. Today, the convergence scenarios are classified in terms of the limiting matrices (i.e. the final possible patterns of the matrices in iteration). This view was first raised by Hoffmann–Parlett [1], and then followed by other researchers [4, 5, 8]. Although this classification has been practically successful in itself, by carefully observing the statement we soon notice an interesting fact that formally it does not mention all the possible limiting matrix patterns (see Section 3 for the detail). This may be partly because the unmentioned pattern is relatively exceptional, and most initial matrices result in either of the mentioned patterns. Mathematically strictly speaking, however, it is desired to fill this gap and cover all the possible patterns.

In this paper, we first show by an example that the unmentioned pattern in fact happens. Then we theoretically show that all the examples belonging to this new pattern enjoy cubic convergence (thus, in other words, we extended the class of cubically convergent matrices). This gives a new "complete" classification, which covers all the possible convergence scenarios. Curiously, it turns out that the new classification is simpler than the original classification.

This paper is organized as follows. After a brief summary of the shifted QR algorithm in Section 2, we summarize the standard classification owing to Hoffmann–Parlett [1] in Section 3. Then in Section 4, we present an example that achieves the unmentioned convergence pattern. Section 5 is the main part where the new classification and its theoretical analysis is given. Section 6 is devoted to the conclusions. For readers' convenience, a global convergence proof of the QR algorithm (in the case of symmetric tridiagonal matrices) is given in Appendix.

2 QR algorithm with Wilkinson's shift

In this section the QR algorithm with Wilkinson's shift and its convergence theorem are summarized.

Let us write a symmetric irreducible tridiagonal matrix as

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

The eigenvalues of T are all distinct [4], which are denoted as $\lambda_1 > \cdots > \lambda_m$ here. The QR algorithm is described as follows.

Algorithm 1 QR algorithm

Initialization: $T^{(0)} := T$

1: **for** $n := 0, 1, \dots$ **do**

2: Choose shift $s^{(n)}$

3: $T^{(n)} - s^{(n)} I = Q^{(n)} R^{(n)}$

4: $T^{(n+1)} := R^{(n)} Q^{(n)} + s^{(n)} I$

5: **end for**

Similarly to (1), tridiagonal elements of $T^{(n)}$ are denoted by

$$T^{(n)} = \begin{pmatrix} \alpha_1^{(n)} & \beta_1^{(n)} & & \\ \beta_1^{(n)} & \alpha_2^{(n)} & \ddots & \\ & \ddots & \ddots & \beta_{m-1}^{(n)} \\ & & \beta_{m-1}^{(n)} & \alpha_m^{(n)} \end{pmatrix}. \tag{2}$$

In line 2 of the above algorithm, one way to determine an efficient shift $s^{(n)}$ is to consider the lower right 2-by-2 submatrix, and pick its eigenvalue closer to $\alpha_m^{(n)}$. This is the so called Wilkinson shift. In this case the global convergence is theoretically guaranteed, and the convergence rate is at least quadratic as the next theorem indicates.

Theorem 1 (Wilkinson [7]). *Suppose the QR algorithm with Wilkinson's shift is applied to an irreducible tridiagonal matrix T. Then we have*

$$\lim_{n\to\infty} \alpha_m^{(n)} = \lambda_l, \qquad \lim_{n\to\infty} |\beta_{m-1}^{(n)}| = 0, \qquad |\beta_{m-1}^{(n+1)}| = \mathrm{O}(|\beta_{m-1}^{(n)}|^2), \tag{3}$$

where λ_l is one of the eigenvalues of T. ■

3 Convergence classification by Hoffmann–Parlett [1, 4]

Despite the above theorem, quite often it is observed that Wilkinson's shift achieves cubic convergence. This phenomena is mathematically described in the next theorem, which states that if the second lower right element $\beta_{m-2}^{(n)}$ converges to 0, then the lower right element $\beta_{m-1}^{(n)}$ enjoys cubic convergence[1].

Theorem 2 (Hoffmann–Parlett [1, 4]). *Suppose the QR algorithm with Wilkinson's shift is applied to an irreducible tridiagonal matrix T. If the lower right 3-by-3 submatrix converges to the limiting matrix*

$$\lim_{n\to\infty} \begin{pmatrix} \alpha_{m-2}^{(n)} & \beta_{m-2}^{(n)} & 0 \\ \beta_{m-2}^{(n)} & \alpha_{m-1}^{(n)} & \beta_{m-1}^{(n)} \\ 0 & \beta_{m-1}^{(n)} & \alpha_m^{(n)} \end{pmatrix} = \begin{pmatrix} * & 0 & 0 \\ 0 & \lambda_k & 0 \\ 0 & 0 & \lambda_l \end{pmatrix}, \tag{4}$$

[1]Here we exclude the case where Wilkinson's shift happens to coincide with an eigenvalue of T at some finite step n. In this case, the QR iteration is immediately terminated, and hence asymptotic convergence analysis is not needed.

where $k \neq l$, *then* $|\beta_{m-1}^{(n+1)}| = \mathrm{O}(|\beta_{m-2}^{(n)}|^2 |\beta_{m-1}^{(n)}|^3)$. *If the lower right 3-by-3 submatrix converges to the limiting matrix*

$$\lim_{n\to\infty} \begin{pmatrix} \alpha_{m-2}^{(n)} & \beta_{m-2}^{(n)} & 0 \\ \beta_{m-2}^{(n)} & \alpha_{m-1}^{(n)} & \beta_{m-1}^{(n)} \\ 0 & \beta_{m-1}^{(n)} & \alpha_m^{(n)} \end{pmatrix} = \begin{pmatrix} * & C & 0 \\ C & \lambda_l & 0 \\ 0 & 0 & \lambda_l \end{pmatrix}, \tag{5}$$

where $C \neq 0$, *then* $|\beta_{m-1}^{(n+1)}| = \mathrm{O}(|\beta_{m-1}^{(n)}|^2)$. ■

In most cases, the limiting matrix satisfies (4), and thus the actual convergence speed is cubic [1, 4, 5, 8]. The possibility of the loss of cubic convergence, i.e., the occurrence of (5), was discussed in several studies (see, for example, [1, 4, 5, 8]), but it is still mathematically open whether or not there in fact exists such a matrix that leads to the case (except the case $m = 3$, for which a rigorous analysis is given in [2]; see Remark 2 below). Clarifying this ambiguity seems considerably difficult, and in the present paper we do not challenge this, taking the same attitude as the other researchers in the literature.

Instead, we here focus on an interesting obvious fact that the patterns (4) and (5) do not cover every possible limiting matrix, at least formally—it excludes the case where the lower two diagonal elements converge to different values, *and* at the same time $\beta_{m-2}^{(n)}$ does not vanish (i.e. $\beta_{m-2}^{(n)} \to C \neq 0$). What happens in that case? Is it excluded since it never happens?

Below we show an answer to these questions; it turns out that such case actually happens (we show an example), and there the convergence is cubic. In other words, there is another class of cubically convergent matrices, which is not mentioned in the previous classification, and not empty.

4 A numerical experiment

Let us apply the QR algorithm with Wilkinson's shift to a 101-by-101 matrix

$$T^{(0)} = \begin{pmatrix} 0 & 100 & & & \\ 100 & \ddots & \ddots & & \\ & \ddots & 0 & 100 & \\ & & 100 & 0 & 1 \\ & & & 1 & 0 \end{pmatrix}, \tag{6}$$

whose eigenvalues are 0, $\pm c_k$ ($k = 1, \ldots, 50$). Also let $T_2^{(n)}$ be the lower right 3-by-3 submatrix of $T^{(n)}$; for $n = 0$ we see

$$T_2^{(0)} = \begin{pmatrix} 0 & 100 & \\ 100 & 0 & 1 \\ & 1 & 0 \end{pmatrix}. \tag{7}$$

Then we find the following convergence behavior of $T_2^{(n)}$ for $n = 3, 4, 5$:

$$T_2^{(3)} = \begin{pmatrix} 1.80 & 2.55 & \\ 2.55 & -1.81 & 2.55 \times 10^{-14} \\ & 2.55 \times 10^{-14} & -1.21 \times 10^{-28} \end{pmatrix},$$

$$T_2^{(4)} = \begin{pmatrix} 1.80 & 2.53 & \\ 2.53 & -1.81 & -1.96 \times 10^{-42} \\ & -1.96 \times 10^{-42} & -7.18 \times 10^{-85} \end{pmatrix},$$

$$T_2^{(5)} = \begin{pmatrix} 1.81 & 2.53 & \\ 2.53 & -1.81 & -8.88 \times 10^{-127} \\ & -8.88 \times 10^{-127} & -1.47 \times 10^{-253} \end{pmatrix}.$$

Thus it tends to

$$\lim_{n\to\infty} \begin{pmatrix} \alpha_{m-2}^{(n)} & \beta_{m-2}^{(n)} & 0 \\ \beta_{m-2}^{(n)} & \alpha_{m-1}^{(n)} & \beta_{m-1}^{(n)} \\ 0 & \beta_{m-1}^{(n)} & \alpha_m^{(n)} \end{pmatrix} = \begin{pmatrix} \lambda_l - D & C & 0 \\ C & \lambda_l + D & 0 \\ 0 & 0 & \lambda_l \end{pmatrix}, \tag{8}$$

where $\lambda_l = 0$, $C \approx 2.5$, $D \approx -1.8$. Obviously this example does not belong to either of (4) and (5). The observed convergence rate of $\beta_{m-1}^{(n)}$ is cubic: $|\beta_{m-1}^{(n+1)}|/|\beta_{m-1}^{(n)}|^3 \approx 0.11$. Despite the long history of the convergence analysis for Wilkinson's shift, no similar example has been pointed out in the research field of numerical linear algebra, as far as the authors know. In this sense, one of the contribution of this paper is the founding of the matrix (8). In the next section, we will prove that the matrix in fact theoretically enjoys cubic convergence, and then propose a new classification reflecting the fact.

5 New classification for convergence rate

In this section we rectify the existing classification so that the case (8) can be covered. The new classification, which is summarized in the next theorem, focuses on the lower right 2-by-2 submatrix, instead of 3-by-3 in the previous classification.

Theorem 3. *Suppose the QR algorithm with Wilkinson's shift is applied to an irreducible tridiagonal matrix T. If the lower right 2-by-2 submatrix converges to the limiting matrix*

$$\lim_{n\to\infty} \begin{pmatrix} \alpha_{m-1}^{(n)} & \beta_{m-1}^{(n)} \\ \beta_{m-1}^{(n)} & \alpha_m^{(n)} \end{pmatrix} = \begin{pmatrix} \lambda_l + D & 0 \\ 0 & \lambda_l \end{pmatrix}, \tag{9}$$

where $D \neq 0$, then $|\beta_{m-1}^{(n+1)}| = \mathrm{O}(|\beta_{m-2}^{(n)}|^2 |\beta_{m-1}^{(n)}|^3)$. If the lower right 2-by-2 submatrix converges to the limiting matrix

$$\lim_{n\to\infty} \begin{pmatrix} \alpha_{m-1}^{(n)} & \beta_{m-1}^{(n)} \\ \beta_{m-1}^{(n)} & \alpha_m^{(n)} \end{pmatrix} = \begin{pmatrix} \lambda_l & 0 \\ 0 & \lambda_l \end{pmatrix}, \tag{10}$$

then $|\beta_{m-1}^{(n+1)}| = \mathrm{O}(|\beta_{m-1}^{(n)}|^2)$.

Proof. We here focus on the first claim, since the second one has been essentially already settled in the literature. In order to reveal the convergence rate, let us estimate the distance between Wilkinson's shift and an eigenvalue of $T^{(n)}$ closer to the shift. To this end, suppose we apply one step of the Jacobi method for the lower right 2-by-2 submatrix of $T^{(n)}$. Then we see the angle $\theta^{(n)}$ of Givens rotation to annihilate $\beta_{m-1}^{(n)}$ satisfies

$$\tan(2\theta^{(n)}) = \frac{2\beta_{m-1}^{(n)}}{\alpha_{m-1}^{(n)} - \alpha_m^{(n)}} \tag{11}$$

from [4, Chapter 9], where $\theta^{(n)}$ is chosen in the interval $[-\pi/4,\ \pi/4]$, and $\theta^{(n)} = \mathrm{O}(\beta_{m-1}^{(n)})$ as $n \to \infty$ in view of the limiting matrix (9). It means that the transformed matrix can be described as

$$\begin{pmatrix} \ddots & \ddots & & & & \\ \ddots & * & * & & & \\ & * & * & * & z^{(n)} & \\ & & * & * & 0 & \\ & & z^{(n)} & 0 & s^{(n)} & \end{pmatrix},$$

where

$$z^{(n)} = \beta_{m-2}^{(n)} \sin(\theta^{(n)}). \tag{12}$$

Note that $s^{(\infty)} = \lambda_l$. Let $\delta = \min_{i \neq l} |\lambda_i - s^{(\infty)}|$. For any $\epsilon > 0$, we see

$$|\lambda_l - s^{(n)}| \leq \frac{|z^{(n)}|^2}{\delta - \epsilon} \tag{13}$$

for all sufficiently large n from the so-called gap theorem [4, Theorem 11.7.1]. Let λ_k be the closest eigenvalue to λ_l. Since the convergence rate of the subdiagonal element for the QR algorithm is the ratio of the eigenvalues, we see

$$|\beta_{m-1}^{(n+1)}| \sim \frac{|\lambda_l - s^{(n)}|}{|\lambda_k - s^{(n)}|} |\beta_{m-1}^{(n)}|$$

as $n \to \infty$. Therefore we obtain

$$\begin{aligned} |\beta_{m-1}^{(n+1)}| &\leq |\beta_{m-1}^{(n)}||z^{(n)}|^2/|\delta - \epsilon||\delta| \\ &= |\beta_{m-1}^{(n)}||\beta_{m-2}^{(n)} \sin(\theta^{(n)})|^2/|\delta - \epsilon||\delta| \\ &= |\beta_{m-1}^{(n)}|^3|\beta_{m-2}^{(n)}|^2/|\alpha_{m-1}^{(n)} - \alpha_m^{(n)}|^2|\delta - \epsilon||\delta| \end{aligned}$$

as $n \to \infty$, by using (13), (12), (11) in turn. □

As noted above, the previous classification did not explicitly considered the case (8). In our experience, this case is relatively rare compared to the previously known cubic case (4), but not extremely rare so that it can be left outside our consideration (see also Remark 2 below). The above new classification is "complete" in the sense that it also covers (8). It is also preferable in that it is quite simple; it suffices to check the lower 2-by-2 matrix. If the two diagonal elements converge to different values, cubic convergence is achieved. Otherwise the rate is at least quadratic, as already proved by several authors.

Remark 1. Strictly speaking, in order for such classifications by limiting matrices to work for *every* initial matrix, it should be also proved that the (3-by-3 or 2-by-2) submatrix in question always tends to a constant matrix (without exhibiting any oscillatory behaviors). Actually this holds true. This fact might have been noticed by the experts in this research field because its proof is almost the same as that by [3, 6] for the unshifted QR algorithm. However, the present authors do not know any reference where the proof for the shifted algorithm is explicitly stated. For the readers' convenience, in the present paper a stronger result stating that *all* of the tridiagonal elements in fact tend to constants is shown in Appendix. ■

Remark 2. The above result is closely related to the work by Leitte–Saldanha–Tomei [2], which considered the case of 3-by-3 matrices, and proved in the language of dynamical systems theory that there exists a 3-by-3 matrix that converges only quadratically by the QR iteration with Wilkinson's shift. More precisely, they considered the matrix

$$\tilde{T} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{14}$$

and regarded the QR iterations with Wilkinson's shift as maps generating a discrete dynamical system in the space of symmetric tridiagonal matrices with the same spectrum as $\tilde{T}$ (see the original paper for the detail). Then they proved that there exists an open neighborhood of $\tilde{T}$ such that (i) the iteration maps a point (matrix) back to the set, (ii) the convergence is strictly quadratic if it tends to $\tilde{T}$ and cubic otherwise, and (iii) the Hausdorff dimension of the set of such initial points that leads to the quadratic convergence is 1.

Although the main topic of [2] is the existence of the quadratic cases, if we view the result from the opposite direction, it is also claiming that there are quite many initial matrices close to $\tilde{T}$ resulting in cubic convergence (recall (iii), which states that the quadratic cases are "very thin" [2]). This strongly suggests that cubic convergence can be observed even when $\beta_{m-2}^{(n)}$ does not tend to 0 (since it should stay around 1). In this way, the work [2] suggests a similar result as above in the case of 3-by-3 matrices, although explicit initial matrix examples are not given. By shrinking the size of the example (6) in the present paper, it is easy to obtain a 3-by-3 example

$$T^{(0)} = \begin{pmatrix} 0 & 100 & \\ 100 & 0 & 1 \\ & 1 & 0 \end{pmatrix}, \tag{15}$$

which actually tends to the form (8) cubically. ■

6 Conclusions and future works

We pointed out a matrix example enjoying cubic convergence, which was not covered in the previous standard classification. It tends to the unmentioned limiting form (8). In order to neatly cover this case, it is more natural to consider the lower 2-by-2 submatrix rather than 3-by-3. Taking into account the fact that all the elements always converge (see Remark 1), we see that the classification by Theorem 3 is "complete" in the sense that it covers all the possible scenarios, starting from any initial matrix.

It is still open, however, whether or not there actually exists a strictly quadratic case for 4-by-4 or larger matrices, i.e., if the limiting case (5) actually occurs. Furthermore, even if it is confirmed, still there remains the possibility that the estimate $|\beta_{m-1}^{(n+1)}| = \mathrm{O}(|\beta_{m-1}^{(n)}|^2)$ is an overestimate, and the actual rate is cubic. These issues are left as future works. (Recall that for 3-by-3 matrices they have been completely settled in [2]; see Remark 2).

Acknowledgments

The authors are grateful to Professor Parlett for his valuable comments and suggestions. The first author is supported by JSPS Grant-in-Aid for Research Activity Start-up.

References

[1] W. Hoffmann and B. N. Parlett, A new proof of global convergence for the tridiagonal QL algorithm, *SIAM Journal on Numerical Analysis*, vol. 15 (1978), pp. 929–937.

[2] R. S. Leitte, N. C. Saldanha and C. Tomei: The asymptotics of Wilkinson's shift: loss of cubic convergence, *Foundations of Computational Mathematics*, vol. 10 (2010), pp. 15–36.

[3] B. N. Parlett: Global convergence of the basic QR algorithm on Hessenberg matrices, *Mathematics of Computation*, vol. 22 (1968), pp. 803–817.

[4] B. N. Parlett: *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, New Jersey, 1980; SIAM, Philadelphia, 1998.

[5] T. Wang: Convergence of the tridiagonal QR algorithm, *Linear Algebra and Its Applications*, vol. 322 (2001), pp. 1–17.

[6] J. H. Wilkinson: *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

[7] J. H. Wilkinson: Global convergence of tridiagonal QR algorithm with origin shifts, *Linear Algebra and Its Applications*, vol. 1 (1968), pp. 409–420.

[8] G. Zhang: On the convergence rate of the QL algorithm with Wilkinson's shift, *Linear Algebra and Its Applications*, vol. 113 (1989), pp. 131–137.

7 Appendix

Here we discuss the global convergence of the QR algorithm; more precisely, we prove that in the case of symmetric tridiagonal matrices all the tridiagonal elements converge for any initial matrix.

Suppose a general shift $s^{(n)}$ satisfying the following conditions:

(i) The shift $s^{(n)}$ converges to a certain eigenvalue $s^{(\infty)} = \lambda_l$;

(ii) $|s^{(n)} - \lambda_l| = \mathrm{o}(c^n)$ for a positive constant $c < 1$.

Note that Wilkinson's shift satisfies the two conditions: (i) has been proved by [5]; then the convergence rate by the shifts $s^{(n)} \to \lambda_l$ is at least quadratic because $|s^{(n)} - \lambda_l| \le |s^{(n)} - \alpha_m^{(n)}| + |\alpha_m^{(n)} - \lambda_l| \le 2|\beta_{m-1}^{(n)}|$ by the Gershgorin's circle theorem and $|\beta_{m-1}^{(n+1)}| = \mathrm{O}(|\beta_{m-1}^{(n)}|^2)$ by Wilkinson's proof, which implies (ii).

In what follows, we show the global convergence for such general shifts. The following convergence proof might have been noticed by the experts in this research field because its proof is almost the same as that by [3, 6] for the unshifted QR algorithm. However, to the best of the authors' knowledge, the proof for the shifted algorithm is not explicitly stated in any reference. For the readers' convenience, we prove it as follows.

Similarly to the discussion in [3] and [6, Chapter 8, §28], let $\tilde{Q}^{(n)}$, $\tilde{R}^{(n)}$ be

$$\begin{aligned} \tilde{Q}^{(n)} &= Q^{(0)} \cdots Q^{(n-1)}, \\ \tilde{R}^{(n)} &= R^{(n-1)} \cdots R^{(0)}. \end{aligned}$$

Then we see

$$(T - s^{(0)} I) \cdots (T - s^{(n-1)} I) = \tilde{Q}^{(n)} \tilde{R}^{(n)}. \tag{16}$$

By the orthogonal matrix $\tilde{Q}^{(n)}$, $T^{(n)}$ is described as

$$T^{(n)} = (\tilde{Q}^{(n)})^{\mathrm{T}} T \tilde{Q}^{(n)}. \tag{17}$$

Let $p(l)$ denote a permutation of the indices l $(l = 1, \ldots, m)$. Then in view of the condition (i) we can we place the shifted eigenvalues in a descending order as

$$|\lambda_{p(1)} - s^{(\infty)}| \geq \cdots \geq |\lambda_{p(m-1)} - s^{(\infty)}| > |\lambda_{p(m)} - s^{(\infty)}| = 0. \tag{18}$$

The last inequality follows from (i) (note that now we are assuming that all the eigenvalues are distinct).

Next, we focus on the eigendecomposition

$$T = X \Lambda X^{\mathrm{T}}, \tag{19}$$

where X is the orthogonal matrix consisting of the eigenvectors and Λ is the diagonal matrix with the eigenvalues: $\mathrm{diag}(\lambda_{p(1)}, \ldots, \lambda_{p(m)})$. Then we see

$$(T - s^{(0)} I) \cdots (T - s^{(n-1)} I) = X \Lambda^{(n)} X^{\mathrm{T}}, \tag{20}$$

where

$$\Lambda^{(n)} = (\Lambda - s^{(0)} I) \cdots (\Lambda - s^{(n-1)} I). \tag{21}$$

Here we apply the LU factorization $X^{\mathrm{T}} = LU$ (note that X^{T} constructed by the normalized eigenvectors of an irreducible tridiagonal matrix is always LU factorizable). It then follows that

$$(T - s^{(0)} I) \cdots (T - s^{(n-1)} I) = X \Lambda^{(n)} L (\Lambda^{(n)})^{-1} \Lambda^{(n)} U. \tag{22}$$

Combining it with (16) we have

$$\tilde{Q}^{(n)} \tilde{R}^{(n)} = X \Lambda^{(n)} L (\Lambda^{(n)})^{-1} \Lambda^{(n)} U. \tag{23}$$

Let $D_{\Lambda^{(n)}}$ be a unitary diagonal matrix

$$\mathrm{diag}((\lambda_{p(1)} - s^{(n)})/|\lambda_{p(1)} - s^{(n)}|, \ldots, (\lambda_{p(m)} - s^{(n)})/|\lambda_{p(m)} - s^{(n)}|). \tag{24}$$

It is easy to see that

$$D_{\Lambda^{(n)}} \Lambda^{(n)} = \mathrm{diag}(|\lambda_{p(1)} - s^{(n)}|, \ldots, |\lambda_{p(m)} - s^{(n)}|). \tag{25}$$

In the right-hand side of (23), we have

$$\Lambda^{(n)} L (\Lambda^{(n)})^{-1} = D_{\Lambda^{(n)}}^{-1} (D_{\Lambda^{(n)}} \Lambda^{(n)} L (D_{\Lambda^{(n)}} \Lambda^{(n)})^{-1}) D_{\Lambda^{(n)}}, \tag{26}$$

and by applying the QR factorization we see

$$D_{\Lambda^{(n)}} \Lambda^{(n)} L (D_{\Lambda^{(n)}} \Lambda^{(n)})^{-1} = P^{(n)} \Gamma^{(n)}, \tag{27}$$

where $P^{(n)}$ is an orthogonal matrix, $\Gamma^{(n)}$ is an upper triangular matrix whose diagonal elements are positive. Let D_U be a unitary diagonal matrix $D_U = \mathrm{diag}(u_{11}/|u_{11}|, \ldots, u_{mm}/|u_{mm}|)$. Then we see

$$\tilde{Q}^{(n)} = XD_{\Lambda^{(n)}}^{-1}P^{(n)}D_U^{-1} \tag{28}$$

$$\tilde{R}^{(n)} = D_U\Gamma^{(n)}D_{\Lambda^{(n)}}\Lambda^{(n)}U \tag{29}$$

from (23), (26), and (27). Therefore, we have

$$T^{(n)} = D_U(P^{(n)})^{\mathrm{T}}\Lambda P^{(n)}D_U^{-1} \tag{30}$$

from (17) and (19).

Since our aim is to prove the convergence of all the elements of $T^{(n)}$, let us discuss the behavior of the orthogonal matrix $P^{(n)}$ as $n \to \infty$. To this end, we focus on (27). The lower left elements are

$$(D_{\Lambda^{(n)}}\Lambda^{(n)}L(D_{\Lambda^{(n)}}\Lambda^{(n)})^{-1})_{ij} = l_{ij}\prod_{l=0}^{n-1}\left|\frac{\lambda_{p(j)} - s^{(l)}}{\lambda_{p(i)} - s^{(l)}}\right| \quad (i > j) \tag{31}$$

from (25). Obviously $\lim_{n\to\infty}(D_{\Lambda^{(n)}}\Lambda^{(n)}L(D_{\Lambda^{(n)}}\Lambda^{(n)})^{-1})_{ij} = 0$, when $|\lambda_{p(i)} - s^{(\infty)}| > |\lambda_{p(j)} - s^{(\infty)}|$. Otherwise, from (18) and the condition (ii), we have

$$\left|\frac{\lambda_{p(j)} - s^{(l)}}{\lambda_{p(i)} - s^{(l)}}\right| = \left|\frac{\lambda_{p(j)} - \lambda_{p(m)} + \lambda_{p(m)} - s^{(l)}}{\lambda_{p(i)} - \lambda_{p(m)} + \lambda_{p(m)} - s^{(l)}}\right| = 1 + \mathrm{o}(c^l). \tag{32}$$

Since a sequence of the size $\mathrm{o}(c^l)$ with $0 < c < 1$ absolutely converges, $(D_{\Lambda^{(n)}}\Lambda^{(n)}L(D_{\Lambda^{(n)}}\Lambda^{(n)})^{-1})_{ij}$ represented by the infinite product (31) is convergent:

$$\lim_{n\to\infty} D_{\Lambda^{(n)}}\Lambda^{(n)}L(D_{\Lambda^{(n)}}\Lambda^{(n)})^{-1} = \tilde{L}. \tag{33}$$

The resulting matrix $\tilde{L}$ is not only unit lower triangular, but also block diagonal with the block sizes at most 2, because the equality in (18) can appear only once thanks to the fact that the eigenvalues are all distinct. Hence, the orthogonal matrix $P^{(n)}$ given by the QR factorization of $D_{\Lambda^{(n)}}\Lambda^{(n)}L(D_{\Lambda^{(n)}}\Lambda^{(n)})^{-1}$ is convergent:

$$\lim_{n\to\infty} P^{(n)} = \tilde{P}, \tag{34}$$

where $\tilde{P}$ is a block diagonal matrix whose block size is at most 2. It then follows that

$$\lim_{n\to\infty} T^{(n)} = D_U\tilde{P}^{\mathrm{T}}\Lambda\tilde{P}D_U^{-1} \tag{35}$$

from (30). Therefore, $T^{(n)}$ converges to a block diagonal matrix whose block size is at most 2.