

## ON THE $QR$ ITERATIONS OF REAL MATRICES

HUAJUN HUANG AND TIN-YAU TAM

ABSTRACT. We answer a question of D. Serre on the  $QR$  iterations of a real matrix with nonreal eigenvalues whose moduli are distinct except for the conjugate pairs. Numerical experiments by MATLAB are performed.

### 1. INTRODUCTION

There are many numerical methods for the computation of the eigenvalues of a given  $A \in GL_n(\mathbb{K})$  with  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$ . One of the most efficient methods is the  $QR$  method [3, p.173-180]. Define a sequence  $\{A_k\}_{k \in \mathbb{N}} \subseteq GL_n(\mathbb{K})$  of matrices with  $A_1 := A$  and  $A_{j+1} := R_j Q_j$  if  $A_j = Q_j R_j$  is the  $QR$  decomposition of  $A_j$ ,  $j = 1, 2, \dots$ . Notice that

$$(1.1) \quad A_{j+1} = Q_j^{-1} A_j Q_j.$$

So the eigenvalues of each  $A_j$  are identical with those of  $A$ , counting multiplicities. One hopes to have some sort of convergence on the sequence  $\{A_k\}_{k \in \mathbb{N}}$  so that the “limit” would provide the eigenvalues of  $A$ . If we write

$$P_k = Q_1 Q_2 \cdots Q_k, \quad U_k = R_k R_{k-1} \cdots R_1,$$

then [3]

$$(1.2) \quad A^k = P_k U_k, \quad Q_k = P_{k-1}^{-1} P_k, \quad R_k = U_k U_{k-1}^{-1},$$

and

$$(1.3) \quad A_k = P_{k-1}^{-1} A P_{k-1} = U_{k-1} A U_{k-1}^{-1}.$$

In Wilkinson’s book [6, p.517-518] one finds the following classical result.

**Theorem 1.1.** *Let  $A \in GL_n(\mathbb{C})$  such that the moduli of the eigenvalues  $\lambda_1, \dots, \lambda_n$  of  $A$  are distinct, that is,*

$$(1.4) \quad |\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| (> 0).$$

*Let  $A = Y^{-1} \text{diag}(\lambda_1, \dots, \lambda_n) Y$ . Assume that  $Y$  admits an  $LU$  decomposition  $Y = LU$ . Then the strictly lower triangular part of  $A_k$  converges to zero and the diagonal part of  $A_k$  converges to  $D := \text{diag}(\lambda_1, \dots, \lambda_n)$ .*

Though Theorem 1.1 is a rather satisfactory result, in many applications one encounters  $A \in GL_n(\mathbb{R})$ . If  $A$  has nonreal eigenvalues, then they occur in complex conjugate pairs and the assumption (1.4) does not hold for  $A$ .

D. Serre [3, p.174] asserts that

---

2000 AMS Mathematics Subject Classification. Primary 15A23, 65F10  
 Key words:  $QR$  iterations, eigenvalues,  $QR$  decomposition,  $LU$  decomposition

“When  $A \in M_n(\mathbb{R})$ , one makes the following assumption. Let  $p$  be the number of real eigenvalues and  $2q$  that of nonreal eigenvalues; then there are  $p + q$  distinct eigenvalue moduli. In that case,  $\{A_k\}_{k \in \mathbb{N}}$  might converge to a block-triangular form, the diagonal blocks being  $2 \times 2$  or  $1 \times 1$ . The limits of the diagonal blocks provide trivially the eigenvalues of  $A$ .”

The assertion has never been proved nor disproved, as pointed out by Serre [3, p.175]. Evidently the above quoted paragraph is interpreted as the strictly lower triangular block part of  $\{A_k\}_{k \in \mathbb{N}}$  converges to zero. Indeed the diagonal blocks of  $\{A_k\}_{k \in \mathbb{N}}$  may not converge, even though the eigenvalues of these diagonal blocks converge to the eigenvalues of  $A$  (see Proposition 4.2).

In Section 2, Theorem 2.1 gives an affirmative answer to the question of Serre under a very mild condition. Namely, if a real matrix  $A = Y^{-1}DY$  has distinct moduli eigenvalues (up to conjugate pairs), where  $D$  is given in (2.2), and  $Y$  admits a certain block  $LU$  decomposition, then the strictly lower triangular block part of  $\{A_k\}_{k \in \mathbb{N}}$  converges to zero, where the diagonal blocks (of  $2 \times 2$  or  $1 \times 1$  forms) provide the eigenvalues of  $A$ .

In Section 3, we exhibit that if  $Y$  ( $A = Y^{-1}DY$ ) does not have the block  $LU$  decomposition as in Theorem 2.1, then the conclusion of Theorem 2.1 may not hold, based on some numerical experiments. In other words, Serre’s assertion is not true for this kind of matrices.

In Section 4 we provide some quantitative analysis for the  $2 \times 2$  case.

In Section 5 we prove that unlike the real case, the complex case still behaves well even if  $Y$  does not admit  $LU$  decomposition as long as (1.4) is satisfied.

## 2. AN ANSWER TO SERRE’S QUESTION

The assumption of Serre on  $A \in GL_n(\mathbb{R})$  amounts to that the eigenvalues of  $A$  have distinct moduli except for the conjugate pairs. It may be interpreted as the real counterpart of (1.4) in Theorem 1.1. By the real Jordan canonical form [2, Theorem 3.4.5, p.152],  $A$  admits the following decomposition

$$(2.1) \quad A = Y^{-1}DY$$

where

$$(2.2) \quad D := \text{diag} (\lambda_1 E_{\theta_1}, \dots, \lambda_m E_{\theta_m}), \quad \lambda_1 > \dots > \lambda_m > 0,$$

and

$$E_{\theta_i=0} := 1, \quad E_{\theta_i=\pi} := -1, \quad E_{\theta_i \in (0, \pi)} := \begin{pmatrix} \cos \theta_i & \sin \theta_i \\ -\sin \theta_i & \cos \theta_i \end{pmatrix}.$$

In general  $Y$  has the Bruhat decomposition  $Y = L\omega U$  where  $L$  is unit lower triangular,  $U$  is upper triangular, and  $\omega$  is a permutation matrix uniquely determined by  $Y$ . If  $Y$  admits a “block  $LU$  decomposition” analogous to that in Theorem 1.1, we have the following result. Since such matrices  $Y$  form a dense subset of  $GL_n(\mathbb{R})$ , a randomly chosen  $A \in GL_n(\mathbb{R})$  almost surely satisfies the above requirements.

**Theorem 2.1.** *Let  $A \in GL_n(\mathbb{R})$  be a matrix such that the eigenvalues of  $A$  have distinct moduli except for the conjugate pairs. With the above notations, let  $\gamma = (\gamma_1, \dots, \gamma_m)$  where  $\gamma_i$  is the size of  $E_{\theta_i}$ ,  $i = 1, \dots, m$ . Let  $[M]_\gamma$  be the block form*

of  $M$  corresponding to the partitions  $\gamma$ . Let

$$t := \max \left\{ \left| \frac{\lambda_2}{\lambda_1} \right|, \dots, \left| \frac{\lambda_m}{\lambda_{m-1}} \right| \right\}.$$

If  $Y = L\omega U$  and  $[\omega]_\gamma$  is block diagonal (for example, if  $\omega$  is the identity matrix), then the strictly lower triangular block part of  $[A_k]_\gamma$  converges to zero in  $O(t^k)$ , and the eigenvalues of the  $i$ th diagonal block of  $[A_k]_\gamma$  converge to the eigenvalues of  $\lambda_i E_{\theta_i}$  in  $O(t^k)$ .

**Proof:** Let  $Y^{-1} = QR$  so that

$$A^k = Y^{-1} D^k Y = Q R D^k L \omega U = Q R (D^k L D^{-k}) D^k \omega U.$$

Denote  $[L]_\gamma = (L_{ij})_{m \times m}$  where the  $(i, j)$  block of  $L$  is  $L_{ij}$  of size  $\gamma_i \times \gamma_j$ . Then

$$[D^k L D^{-k}]_\gamma = \left( \begin{array}{cc} \left( \frac{\lambda_i}{\lambda_j} \right)^k & E_{\theta_i}^k L_{ij} E_{\theta_j}^{-k} \\ & \end{array} \right)_{m \times m}.$$

Let

$$D_0 := \text{diag}[L]_\gamma = \begin{pmatrix} L_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & L_{mm} \end{pmatrix},$$

where  $\text{diag}[L]_\gamma$  denotes the block diagonal part of  $[L]_\gamma$ . Denote

$$D_k := D^k D_0 D^{-k} = \begin{pmatrix} E_{\theta_1}^k L_{11} E_{\theta_1}^{-k} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & E_{\theta_m}^k L_{mm} E_{\theta_m}^{-k} \end{pmatrix}, \quad k = 1, 2, \dots$$

Using  $L_{ij} = 0$  for  $i < j$  and  $\|E_{\theta_i}\| = 1$  for all  $i$ , where  $\|\cdot\|$  is the spectral norm,

$$D^k L D^{-k} = (I_n + O(t^k)) D_k.$$

So we have

$$\begin{aligned} A^k &= QR(I_n + O(t^k)) D_k D^k \omega U \\ &= Q(I_n + O(t^k)) R D^k D_0 \omega U \\ &= Q O_k T_k R D^k D_0 \omega U. \end{aligned}$$

Here  $O_k T_k$  is the  $QR$  decomposition of the last  $I_n + O(t^k)$ . By the Gram-Schmidt process one has

$$(2.3) \quad O_k = I_n + O(t^k), \quad T_k = I_n + O(t^k).$$

Since  $[T_k R D^k D_0 \omega U]_\gamma$  is a block upper triangular matrix, its  $Q$ -component  $C_k$  in the  $QR$  decomposition is a block diagonal matrix according to  $\gamma$ . So the  $QR$  decomposition of  $A^k$  is

$$A^k = P_k U_k = (Q O_k C_k) (C_k^{-1} T_k R D^k D_0 \omega U).$$

Hence by the uniqueness of the  $QR$  decomposition

$$P_k = Q O_k C_k, \quad U_k = C_k^{-1} T_k R D^k D_0 \omega U.$$

Therefore, by (2.3)

$$\begin{aligned}
(2.4) \quad A_k &= Q_k R_k = P_{k-1}^{-1} P_k U_k U_{k-1}^{-1} \\
&= C_{k-1}^{-1} O_{k-1}^{-1} O_k T_k R D R^{-1} T_{k-1}^{-1} C_{k-1} \\
&= C_{k-1}^{-1} R D R^{-1} C_{k-1} + O(t^k).
\end{aligned}$$

Because  $C_{k-1}^{-1} R D R^{-1} C_{k-1}$  is block upper triangular, the entries of the strictly lower triangular blocks of  $A_k$  approach zero in  $O(t^k)$ . Moreover, by block multiplication the  $i$ th diagonal block of  $C_{k-1}^{-1} R D R^{-1} C_{k-1}$  is similar to that of  $D$ , namely  $\lambda_i E_{\theta_i}$ . So the eigenvalues of the  $i$ th diagonal block of  $[A_k]_\gamma$  approach those of  $\lambda_i E_{\theta_i}$  in  $O(t^k)$ .  $\square$

Numerical experiments demonstrate the convergence rate in Theorem 2.1.

From the computational point of view, the assumption that  $[\omega]_\gamma$  is in block diagonal form does not impose any difficulty:  $A$  will first be reduced to an Hessenberg form to achieve drastic cost reduction [3, p.176]. Thus we may assume that  $A \in GL_n(\mathbb{R})$  is in irreducible (nonreduced) Hessenberg form. Those nonsingular  $Y$  for which  $A = Y^{-1} D Y$  would have the required  $L\omega U$  decomposition in Theorem 2.1, according to the following result.

**Proposition 2.2.** *Suppose that  $A \in GL_n(\mathbb{R})$  in Theorem 2.1 is in irreducible Hessenberg form. Then for any  $Y \in GL_n(\mathbb{R})$  such that  $A = Y^{-1} D Y$ , it has the decomposition  $Y = L\omega U$ , where  $[\omega]_\gamma$  is in diagonal block form, and  $D$  is given in (2.2).*

**Proof:** For any  $\theta \in \mathbb{R}$ , if  $P := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$ , then

$$\text{diag}(e^{i\theta}, e^{-i\theta}) = P \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} P^{-1}.$$

Let  $S \in GL_n(\mathbb{C})$  be in block diagonal form such that the  $2 \times 2$  diagonal blocks of  $[S]_\gamma$  are  $P$  and the  $1 \times 1$  blocks are 1, according to the partition  $\gamma$ . Then  $A = Y^{-1} S^{-1} \tilde{D} S Y$  where  $\tilde{D}$  is a diagonal matrix such that the diagonal blocks of  $[\tilde{D}]_\gamma$  are either  $\pm \lambda_j$  or  $\lambda_j \text{diag}(e^{i\theta_j}, e^{-i\theta_j})$ . We claim that the matrix  $Z = S Y$  admits  $LU$  decomposition and the argument follows from [3, p.179] (there are some typos in the proof). First notice that the rows of  $Z$  are left eigenvectors of  $A$ , that is, if  $z_1, \dots, z_n$  denote the rows of  $Z$ , then  $z_j A = \mu z_j$ ,  $j = 1, \dots, n$ ,  $\mu = \pm \lambda_j$  or  $\lambda_j e^{\pm i\theta_j}$  since  $Z A = \tilde{D} Z$ . Then  $\{z_1^*, \dots, z_q^*\}^\perp$  is invariant under  $A$  and one has  $\{z_1^*, \dots, z_q^*\}^\perp \oplus \{e_1, \dots, e_q\} = \mathbb{C}^n$ . In other words,  $\det(z_j e_k)_{1 \leq j, k \leq q} \neq 0$ . In other words, the leading principal minors of  $Z$  of order  $q$  is nonzero. So  $Z = S Y$  admits an  $LU$  decomposition. Thus  $S Y = L U$  for some unit lower triangular matrix  $L$  and upper triangular matrix  $U$ . Then  $Y = S^{-1} L U$ . Now the matrix  $S^{-1} L$  is in lower triangular block form with diagonal blocks  $1 \times 1$  or  $2 \times 2$ . Applying Gaussian elimination on  $S^{-1} L$ , one has  $Y = L' \omega U'$  where  $L'$  is (real) unit lower triangular,  $U'$  is (real) upper triangular and  $\omega$  is a diagonal block permutation matrix corresponding to the partition  $\gamma$ . The permutation matrix  $\omega$  is unique.  $\square$

In general the strictly lower triangular part of the (real) sequence  $\{A_k\}_{k \in \mathbb{N}}$  does not converge to zero (Compare [2, p.114]).

**Proposition 2.3.** *Suppose that  $A \in M_n(\mathbb{R})$  has nonreal eigenvalues. Then the strictly lower triangular part of  $\{A_k\}_{k \in \mathbb{N}}$  does not converge to zero.*

**Proof:** The sequence  $\{A_k\}_{k \in \mathbb{N}}$  is contained in the compact set

$$\{X \in M_n(\mathbb{R}) : \|X\|_F = \|A\|_F\},$$

where  $\|A\|_F = (\text{tr } A^*A)^{1/2}$  denotes the Frobenius norm of  $A$ . So there is a convergent subsequence  $\{A_{k_i}\}_{i \in \mathbb{N}}$ . If the strictly lower triangular part of the sequence  $\{A_k\}_{k \in \mathbb{N}}$  converged to zero, then the subsequence would converge to a real upper triangular matrix  $U$ . By the continuity of the eigenvalues (counting multiplicities) [3, p.44], the eigenvalues of  $A$  would be the diagonal entries of  $U$  and would be real, a contradiction.  $\square$

The argument in the above proof works for real singular matrices having nonreal eigenvalues as well.

### 3. NUMERICAL EXPERIMENTS

We now discuss some numerical experiments which show that the conclusion of Theorem 2.1 may not hold if the condition on  $Y$  in Theorem 2.1 is not satisfied. Let

$$A = Y^{-1} \begin{pmatrix} a \cos c & a \sin c & 0 & 0 \\ -a \sin c & a \cos c & 0 & 0 \\ 0 & 0 & b \cos d & b \sin d \\ 0 & 0 & -b \sin d & b \cos d \end{pmatrix} Y,$$

where

$$Y = L\omega U,$$

and

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad \omega = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad U = I_4.$$

Clearly the condition of Theorem 2.1 is not satisfied for  $Y$ . With  $a = 2, b = 1/2$ , numerically we have the following pattern convergence (not actual convergence) of the corresponding matrices. We use the formula in (1.3)  $A_k = P_{k-1}^{-1}AP_{k-1}$  instead of  $A_k = R_{k-1}Q_{k-1}$  to compute  $A_k$  via MAPLE and MATLAB.

- (1) If  $c = 2, d = 1$  (the eigenvalues of  $A$  occur as two distinct complex conjugate pairs),

$$A_k \rightarrow \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \end{pmatrix}, \quad Q_k \rightarrow \begin{pmatrix} * & * & * & 0 \\ * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \end{pmatrix}, \quad P_k \rightarrow \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \end{pmatrix}.$$

- (2) If  $c = 2, d = \pi$  ( $-1/2$  is a double eigenvalue of  $A$ ),

$$A_k \rightarrow \begin{pmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & -1/2 \end{pmatrix},$$

$$Q_k \rightarrow \begin{pmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad P_k \rightarrow \begin{pmatrix} * & * & * & 0 \\ * & * & 0 & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & (-1)^k \end{pmatrix}.$$

(3) If  $c = \pi, d = 1$  ( $-2$  is a double eigenvalue of  $A$ ),

$$A_k \rightarrow \begin{pmatrix} -2 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix},$$

$$Q_k \rightarrow \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix}, \quad P_k \rightarrow \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & 0 & * & * \\ 0 & * & * & * \end{pmatrix}.$$

(4) If  $c = 2, d = \pi/2$  (the eigenvalues of  $A$  occur as two distinct complex conjugate pairs),

$$Q_k \rightarrow \begin{pmatrix} * & * & * & 0 \\ * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \end{pmatrix}.$$

Then for all  $k \in \mathbb{N}$ ,

$$A_{2k} \rightarrow \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \end{pmatrix}, \quad P_{2k} \rightarrow \begin{pmatrix} * & * & * & 0 \\ * & * & 0 & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & (-1)^k \end{pmatrix},$$

and

$$A_{2k-1} \rightarrow \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & 0 \\ 0 & * & * & * \end{pmatrix}, \quad P_{2k-1} \rightarrow \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \end{pmatrix}.$$

(5) If  $c = d = \pi/2$ , then for all  $k \in \mathbb{N}$ ,

$$A_{2k} \rightarrow \begin{pmatrix} 0.0000 & -0.0928 & 3.6270 & 4.4725 \\ 0.3714 & -0.1552 & -0.0527 & 1.7557 \\ -1.0933 & 0.3436 & 0.1167 & 0.6999 \\ -0.0000 & -0.1262 & -0.0429 & 0.0385 \end{pmatrix}$$

and

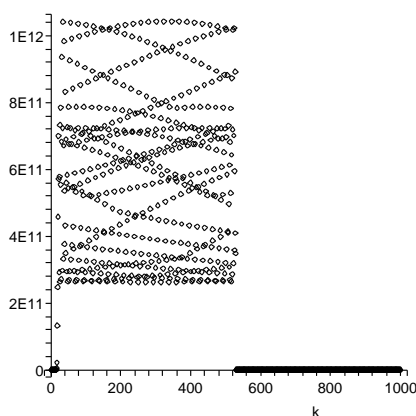
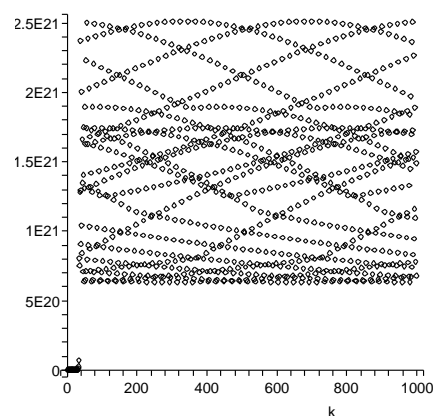
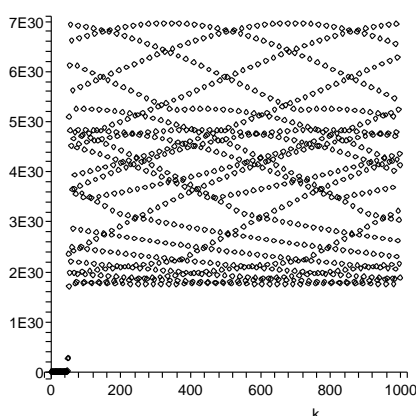
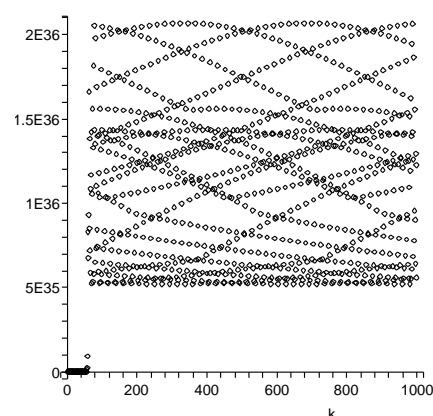
$$A_{2k-1} \rightarrow \begin{pmatrix} -0.0000 & -0.5000 & 1.0607 & -0.3536 \\ 2.0000 & 2.5000 & 1.7678 & 0.3536 \\ -2.8284 & -2.8284 & -2.0000 & 0.0000 \\ -0.0000 & -1.4142 & -1.0000 & -0.5000 \end{pmatrix}.$$

In the above cases, no desired convergence (in the fashion of Theorem 2.1) occurs for the lower triangular block part of  $A_k$ .

We also used  $A_k = R_{k-1}Q_{k-1}$  to compute  $A_k$ . The computed lower triangular block part of  $A_k$  tends to zero. Probably the roundoff errors perturb  $A$  so that the computed  $Y$  has block  $LU$  decomposition in the computational process. Denote

$L_k$  to be the maximal entry in module of the lower left  $2 \times 2$  block of  $A_k$ . The convergence rate of  $L_k$  to 0 is exactly the convergence rate of  $A_k$  to the block upper triangular form. Denote  $c(k) := L_k/t^k$ , where  $t = |\frac{\lambda_2}{\lambda_1}| = \frac{1}{4}$ . When  $A$  meets the conditions in Theorem 2.1, we know  $c(k) \leq M$ ,  $k = 1, 2, \dots$ , for some constant  $M$  depending on  $A$  alone. However, for the above five cases in which  $A$  does not satisfy the conditions in Theorem 2.1, numerical experiments show that we still have  $c(k) \leq M$ , where  $M$  is determined by  $A$  and the digit number used in floating-point computations.

We apply  $A_k = R_{k-1}Q_{k-1}$  in MAPLE to compute  $A_k$  for the first case ( $c = 2$ ,  $d = 1$ ), using 10, 20, 30, 35-digit number floating-point arithmetic, respectively. Then we plot  $c(k)$  against  $k$  for  $1 \leq k \leq 1000$  as follow:

Plot of  $c(k)$  by 10-digit computationPlot of  $c(k)$  by 20-digit computationPlot of  $c(k)$  by 30-digit computationPlot of  $c(k)$  by 35-digit computation

The plots of  $c(k)$  display similar pattern in different floating point precisions. Roughly speaking, when using  $n$ -digit floating-point arithmetic, the upper bound  $M$  of the computed  $c(k)$  is around the scale  $10^n$ . Similar phenomenon holds for the other four cases.

4. ANALYSIS OF THE  $2 \times 2$  REAL CASE WITH NONREAL EIGENVALUES

In Theorem 2.1, we see that the  $QR$  iterations for almost all real matrices converge to a block upper triangular form with  $2 \times 2$  or  $1 \times 1$  diagonal blocks. Thus it is important to study the  $2 \times 2$  real matrix with nonreal eigenvalues in a quantitative fashion.

**Proposition 4.1.** *Suppose  $A \in GL_2(\mathbb{R})$  has nonreal eigenvalues. Let*

$$\frac{A}{\sqrt{\det A}} = Y^{-1} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} Y$$

where

$$Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \in SL_2(\mathbb{R}), \quad \theta \in (0, 2\pi) \setminus \{\pi\}.$$

Denote

$$\begin{aligned} u &:= y_{11}y_{12} + y_{21}y_{22}, \\ v &:= y_{11}^2 + y_{21}^2, \\ r &:= \sqrt{\frac{u^2 + v^2 + 1}{2}} + \sqrt{\left(\frac{u^2 + v^2 + 1}{2}\right)^2 - v^2}. \end{aligned}$$

Then the modulus of the  $(2, 1)$  entry  $c_k$  of  $A_k$  satisfies

$$(4.1) \quad \frac{sv}{r^2} |\sin \theta| \leq |c_k| \leq \min\left\{\frac{sr^2}{v} |\sin \theta|, \|A\|\right\},$$

where  $s := \sqrt{\det A}$  and  $\|A\|$  is the spectral norm of  $A$ .

**Proof:** The singular values of  $A$  and  $A_k$  are the same and thus the entries of  $A_k$  are bounded above by  $\|A\|$ . Notice

$$\begin{aligned} A^k/s^k &= Y^{-1} \begin{pmatrix} \cos k\theta & \sin k\theta \\ -\sin k\theta & \cos k\theta \end{pmatrix} Y \\ &= \begin{pmatrix} y_{22} & -y_{12} \\ -y_{21} & y_{11} \end{pmatrix} \begin{pmatrix} \cos k\theta & \sin k\theta \\ -\sin k\theta & \cos k\theta \end{pmatrix} \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \\ (4.2) \quad &= \begin{pmatrix} \cos k\theta + u \sin k\theta & * \\ -v \sin k\theta & * \end{pmatrix}. \end{aligned}$$

Let  $A^k = P_k U_k = P_k \begin{pmatrix} a_k & * \\ 0 & 1/a_k \end{pmatrix} s^k$ . By the Gram-Schmidt process,

$$\begin{aligned} a_k &= \sqrt{(\cos k\theta + u \sin k\theta)^2 + v^2 (\sin k\theta)^2} \\ &= \sqrt{\frac{u^2 + v^2 + 1}{2} - \frac{u^2 + v^2 - 1}{2} \cos 2k\theta + u \sin 2k\theta} \\ (4.3) \quad &= \sqrt{\frac{u^2 + v^2 + 1}{2} + \sqrt{\left(\frac{u^2 + v^2 - 1}{2}\right)^2 + u^2} \cos(2k\theta + \zeta)}, \end{aligned}$$

where  $\zeta$  is a constant. Since

$$\left(\frac{u^2 + v^2 - 1}{2}\right)^2 + u^2 = \left(\frac{u^2 + v^2 + 1}{2}\right)^2 - v^2,$$

$$\sqrt{\frac{u^2 + v^2 + 1}{2} - \sqrt{\left(\frac{u^2 + v^2 + 1}{2}\right)^2 - v^2}} \leq a_k \leq \sqrt{\frac{u^2 + v^2 + 1}{2} + \sqrt{\left(\frac{u^2 + v^2 + 1}{2}\right)^2 - v^2}}.$$

In other words,

$$(4.4) \quad \frac{v}{r} \leq a_k \leq r.$$

On the other hand, from (2.4)

$$A_k = P_{k-1}^{-1} P_k U_k U_{k-1}^{-1} = P_{k-1}^{-1} A^{k-1} A U_{k-1}^{-1} = U_{k-1} A U_{k-1}^{-1},$$

and

$$(4.5) \quad U_{k-1} = \begin{pmatrix} a_{k-1} & * \\ 0 & 1/a_{k-1} \end{pmatrix} s^{k-1}$$

so that

$$(4.6) \quad \begin{aligned} A_k &= \begin{pmatrix} * & * \\ 0 & 1/a_{k-1} \end{pmatrix} s^k \begin{pmatrix} * & * \\ -v \sin \theta & * \end{pmatrix} s^{-k+1} \begin{pmatrix} 1/a_{k-1} & * \\ 0 & * \end{pmatrix} \\ &= s \begin{pmatrix} * & * \\ -v \sin \theta / a_{k-1}^2 & * \end{pmatrix}. \end{aligned}$$

By (4.4) and (4.6), the modulus of the  $(2, 1)$  entry  $c_k$  of  $A_k$  is bounded by

$$\frac{sv}{r^2} |\sin \theta| \leq |c_k| = | -sv \sin \theta / a_{k-1}^2 | \leq \frac{sr^2}{v} |\sin \theta|.$$

This completes the proof of (4.1).  $\square$

Now we are able to study the convergence of the  $QR$  iterations of the matrix  $A \in GL_2(\mathbb{R})$ . It is sufficient to consider  $A \in SL_2(\mathbb{R})$ .

**Proposition 4.2.**

- (1) *Suppose  $A \in SL_2(\mathbb{R})$  has nonreal eigenvalues. Then  $A_k$  converges if and only if  $A$  is an orthogonal matrix. In this case,  $Q_k = A$ ,  $R_k = I_2$ ,  $k = 1, 2, \dots$*
- (2) *If  $A \in SL_2(\mathbb{R})$  has nonreal eigenvalues and is not an orthogonal matrix, then each of the sequences  $\{A_k\}_{k \in \mathbb{N}}$ ,  $\{P_k\}_{k \in \mathbb{N}}$ ,  $\{U_k\}_{k \in \mathbb{N}}$ ,  $\{Q_k\}_{k \in \mathbb{N}}$ ,  $\{R_k\}_{k \in \mathbb{N}}$  is bounded below and above but not convergent.*

**Proof:** We adopt the notations from Proposition 4.1.

(1) From (4.6), if  $A_k$  converges, then  $c_k$  has to converge. Then by (4.3), we have two possibilities:

- (a)  $(\frac{u^2+v^2-1}{2})^2 + u^2 = 0$ , that is,  $u = 0$  and  $v = 1$ . So by the definitions of  $u$  and  $v$ , the matrix  $Y$  is orthogonal and thus  $A$  is an orthogonal matrix.
- (b)  $\theta = \pi/2$  or  $3\pi/2$ , and  $\cos \zeta = -\frac{u^2+v^2-1}{2} / \sqrt{(\frac{u^2+v^2-1}{2})^2 + u^2} = 0$ . So  $u^2 + v^2 = 1$ , and  $a_k = 1$  by (4.3). We have

$$A = P_1 U_1 = \begin{pmatrix} \cos \eta & \sin \eta \\ -\sin \eta & \cos \eta \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix},$$

for some  $t \in \mathbb{R}$  and  $\eta \in (0, 2\pi) \setminus \{\pi\}$ . If  $t = 0$  then  $A$  is an orthogonal matrix. If  $t \neq 0$  we have

$$\begin{aligned} A_2 &= \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \eta & \sin \eta \\ -\sin \eta & \cos \eta \end{pmatrix} \\ &= \begin{pmatrix} \cos \eta - t \sin \eta & \sin \eta + t \cos \eta \\ -\sin \eta & \cos \eta \end{pmatrix}. \end{aligned}$$

So

$$\begin{aligned} a_2 &= \sqrt{(\cos \eta - t \sin \eta)^2 + (-\sin \eta)^2} \\ &= \sqrt{1 - 2t \cos \eta \sin \eta + t^2 \sin^2 \eta} = 1. \end{aligned}$$

Hence  $t = 2 \cos \eta / \sin \eta$ . In such situation, we have  $A_1 = A_3 = \dots$  and  $A_2 = A_4 = \dots$ . Moreover,  $A_1 = A_2$  if and only if  $\cos \eta = 0$ , contradict with  $t \neq 0$ .

The converse is obviously true.

(2) By (1.2) and (4.5)

$$R_k = U_k U_{k-1}^{-1} = \begin{pmatrix} a_k/a_{k-1} & * \\ 0 & a_{k-1}/a_k \end{pmatrix},$$

since  $s := \sqrt{\det A} = 1$ . By the first part  $A_k$  does not converge. So in (4.3)  $(\frac{u^2+v^2-1}{2})^2 + u^2 \neq 0$ . Thus  $a_k/a_{k-1}$  has finite positive upper bound and lower bound but it does not converge. Thus the entries of  $R_k$  are bound above and below in absolute value but not convergent. Now

$$Q_k = A_k R_k^{-1} = \begin{pmatrix} * & * \\ -v \sin \theta / a_{k-1}^2 & * \end{pmatrix} \begin{pmatrix} a_{k-1}/a_k & * \\ 0 & * \end{pmatrix} = \begin{pmatrix} * & * \\ -(v/a_{k-1} a_k) \sin \theta & * \end{pmatrix}.$$

If  $a_{k-1} a_k$  does not converge, then  $Q_k$  does not converge. If  $a_{k-1} a_k$  converges, then  $A$  belongs to (1)(b) and thus  $Q_k$  does not converge. Neither  $P_k$  nor  $U_k$  converges since  $A_k$  does not converge and (1.3).  $\square$

## 5. SOME REMARKS ON THEOREM 1.1

Theorem 1.1 does not say that  $\{A_k\}_{k \in \mathbb{N}}$  converges to an upper triangular matrix. In fact it is pointed out in [3, p.178], when the eigenvalues of  $A$  have distinct arguments, the sequence  $\{A_k\}_{k \in \mathbb{N}}$  does not converge, in contrast to an incorrect assertion in [2, p.114].

In general  $Y$  in Theorem 1.1 may not admit  $LU$  decomposition. Instead, it has the Bruhat decomposition  $Y = L\omega U$  for some permutation matrix  $\omega \neq I_n$ . However, this will not cause any trouble based on two observations. First the set of nonsingular matrices without  $LU$  decompositions is of measure zero in  $GL_n(\mathbb{C})$  [1, p.407]. So a randomly chosen  $A \in GL_n(\mathbb{C})$  almost surely satisfies the conditions in Theorem 1.1. Secondly, in practice a preliminary reduction of  $A \in GL_n(\mathbb{C})$  to an Hessenberg form drastically reduces the cost of each  $QR$  step [3, p.176]. So  $A$  will first be turned into a Hessenberg form [3, p.169-171, p.175-176], and thus we may assume that  $A$  is in irreducible (nonreduced) Hessenberg form. Those nonsingular  $Y$  satisfying  $A = Y^{-1}DY$  have  $LU$  decomposition [3, p.179].

Nevertheless, for the general case  $Y = L\omega U$ , the diagonal part of  $\{A_k\}_{k \in \mathbb{N}}$  converges to  $D_\omega := \omega^{-1}D\omega$  and the strictly lower triangular part of  $A_k$  converges to zero, which are parts of the statements of Theorem 5.1. Theorem 5.1 is obviously a generalization of Theorem 1.1.

We now introduce some notations. Given an  $n \times n$  permutation  $\omega$ , we denote the permutation on  $\{1, \dots, n\}$  by the same notation  $\omega$  such that  $\omega e_j = e_{\omega(j)}$ ,  $j = 1, \dots, n$ , where  $\{e_1, \dots, e_n\}$  is the standard basis of  $\mathbb{R}^n$ . We write  $O(t^k)$  for a matrix whose entries are less than or equal to  $C|t|^k$  in absolute value for some constant  $C > 0$ .

**Theorem 5.1.** *Let  $A \in GL_n(\mathbb{C})$  such that the moduli of the eigenvalues  $\lambda_1, \dots, \lambda_n$  of  $A$  are distinct, that is,*

$$(5.1) \quad |\lambda_1| > |\lambda_2| > \dots > |\lambda_n| (> 0).$$

Let  $A = Y^{-1}DY$  where

$$D := \text{diag}(\lambda_1, \dots, \lambda_n).$$

Let  $\omega$  be the permutation matrix uniquely determined by  $Y = L\omega U$ , where  $L$  is unit lower triangular and  $U$  is upper triangular. Denote

$$\begin{aligned} H &:= \text{diag}\left(\frac{u_{11}}{|u_{11}|}, \dots, \frac{u_{nn}}{|u_{nn}|}\right), \\ D_\omega &:= \text{diag}(\lambda_{\omega(1)}, \dots, \lambda_{\omega(n)}), \\ C_\omega &:= \text{diag}\left(\frac{\lambda_{\omega(1)}}{|\lambda_{\omega(1)}|}, \dots, \frac{\lambda_{\omega(n)}}{|\lambda_{\omega(n)}|}\right). \end{aligned}$$

Then

$$C_\omega^{k-1}A_kC_\omega^{-k+1} = H^{-1}RD_\omega R^{-1}H + O(t^k), \quad Q_k = C_\omega + O(t^k),$$

where

$$t := \max\left\{\left|\frac{\lambda_2}{\lambda_1}\right|, \dots, \left|\frac{\lambda_n}{\lambda_{n-1}}\right|\right\} < 1,$$

and  $Y^{-1}\omega = QR$  is the QR decomposition of  $Y^{-1}\omega$ . In particular

- (1)  $\lim_{k \rightarrow \infty} C_\omega^{k-1}A_kC_\omega^{-k+1} = \lim_{k \rightarrow \infty} C_\omega^k R_k C_\omega^{-k+1} = H^{-1}RD_\omega R^{-1}H.$
- (2)  $\lim_{k \rightarrow \infty} Q_k = C_\omega.$
- (3) The strictly lower triangular part of  $A_k$  converges to zero in  $O(t^k)$ .
- (4) The diagonal part of  $A_k$  converges to  $D_\omega$  in  $O(t^k)$ .

**Proof:** Under the assumption,

$$A^k = Y^{-1}D^kY.$$

Let  $Y = L\omega U$  where  $L$  is unit lower triangular,  $U$  is upper triangular, and  $\omega$  is a permutation matrix uniquely determined by  $Y$ . Let  $Y^{-1}\omega = QR$  be the QR decomposition of  $Y^{-1}\omega$ . Then

$$A^k = QR\omega^{-1}D^kL\omega U = QR\omega^{-1}(D^kLD^{-k})D^k\omega U.$$

Notice that the unit lower triangular matrix  $D^kLD^{-k} = I_n + O(t^k)$  since the  $(i, j)$  entry of  $D^kLD^{-k}$  where  $i > j$  is  $\ell_{ij}(\lambda_i/\lambda_j)^k$  whose absolute value is less than or equal to  $M|t|^k$ , where  $M = \max_{1 \leq j < i \leq n} |\ell_{ij}|$ . Hence

$$\begin{aligned} A^k &= QR\omega^{-1}(I_n + O(t^k))D^k\omega U \\ &= Q(I_n + O(t^k))R\omega^{-1}D^k\omega U \\ &= QO_kT_kRD_\omega^kU \end{aligned}$$

where  $O_kT_k$  is the QR decomposition of the last  $I_n + O(t^k)$ . By the Gram-Schmidt process we have  $O_k = I_n + O(t^k)$  and  $T_k = I_n + O(t^k)$ . So

$$A^k = P_kU_k = (QO_kHC_\omega^k)(C_\omega^{-k}H^{-1}T_kRD_\omega^kU)$$

is the  $QR$  decomposition of  $A^k$  since  $C_\omega$  and  $H$  are diagonal unitary. By the uniqueness of the  $QR$  decomposition,

$$\begin{aligned} P_k &= QO_kHC_\omega^k = QHC_\omega^k + O(t^k), \\ U_k &= C_\omega^{-k}H^{-1}T_kRD_\omega^kU = C_\omega^{-k}H^{-1}RD_\omega^kU + O(t^k). \end{aligned}$$

By (1.2)

$$(5.2) \quad Q_k = P_{k-1}^{-1}P_k = C_\omega + O(t^k),$$

$$(5.3) \quad R_k = U_kU_{k-1}^{-1} = C_\omega^{-k}H^{-1}RD_\omega R^{-1}HC_\omega^{k-1} + O(t^k),$$

and

$$(5.4) \quad A_k = Q_kR_k = C_\omega^{-k+1}H^{-1}RD_\omega R^{-1}HC_\omega^{k-1} + O(t^k).$$

Since  $C_\omega^{k-1}$  is diagonal unitary,

$$(5.5) \quad C_\omega^{k-1}A_kC_\omega^{-k+1} = H^{-1}RD_\omega R^{-1}H + O(t^k).$$

Conclusion (2) follows from (5.2), (1) from (5.5). From (5.4) the diagonal part of  $A_k$  is

$$\text{diag } A_k = \text{diag } [C_\omega^{-k+1}H^{-1}RD_\omega R^{-1}HC_\omega^{k-1} + O(t^k)] = D_\omega + O(t^k)$$

and (4) follows immediately. Similarly (3) follows from (5.4).  $\square$

**Corollary 5.2.** *The following statements are equivalent.*

- (1) *The sequence  $\{A_k\}_{k \in \mathbb{N}}$  converges.*
- (2) *The sequence  $\{R_k\}_{k \in \mathbb{N}}$  converges.*
- (3) *The arguments of  $\lambda_{\omega(i)}$  and  $\lambda_{\omega(j)}$  are equal whenever the  $(i, j)$  entry of  $RD_\omega R^{-1}$  is nonzero for  $i < j$ .*

**Proof:** (1)  $\Leftrightarrow$  (2) follows from Theorem 5.1 (1) and (2). From (5.4)

$$A_k = H^{-1}C_\omega^{-k+1}RD_\omega R^{-1}C_\omega^{k-1}H + O(t^k).$$

So  $A_k$  converges if and only if  $C_\omega^{-k+1}RD_\omega R^{-1}C_\omega^{k-1}$  converges. Thus (1) and (3) are equivalent.  $\square$

Our proof of Theorem 5.1 is a modification of the proof in [3, Theorem 10.2.1]. Part of the theorem has been discussed in [6, p.519-520] wherein the proof relies on a very careful observation of the pivoting procedure.

**Acknowledgement:** The authors are thankful to the referee's helpful comments.

## REFERENCES

- [1] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, New York, 1978.
- [2] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, 1985.
- [3] D. Serre, *Matrices: Theory and Applications*, Springer, New York, 2002.
- [4] G.W. Stewart, *Introduction to Matrix Computations*, Academic Press, New York, 1973.
- [5] D. Watkins, Understanding the  $QR$  algorithm, *SIAM Rev.* **24** (1982), 427-440.
- [6] J.H. Wilkinson, *The Algebraic Eigenvalues Problem*, Oxford Science Publications, Oxford, 1965.

DEPARTMENT OF MATHEMATICS AND STATISTICS, AUBURN UNIVERSITY, AL 36849-5310, USA  
E-mail address: [huanghu@auburn.edu](mailto:huanghu@auburn.edu), [tamtiny@auburn.edu](mailto:tamtiny@auburn.edu)