JOURNAL OF THE
CHUNGCHEONG MATHEMATICAL SOCIETY
Volume 14, No.2, December 2001

ON THE STABILITY OF THE HOUSEHOLDER QR FACTORIZATION

YUNJUNG AN*, SEYOUNG OH** AND SEIYOUNG CHUNG***

ABSTRACT. A new stability analysis of QR factorizations using the two types of Householder reflections is presented. It shows that the use of the common type(the first) in QR factorizations with pivoting for size by row and column exchanges leas to rowwise more stability, although the second type has been used in Householder transformation for the better purpose. A row-oriented backward error bound for the second type Householder QR factorizations with column pivoting is drived.

1. Introduction

Since Householder transformation was introduced in 1932 by Turnbull and Aitken, it has been proved that Householder transformation, one of orthogonal(unitary) transformations, dose not worsen the condition or stability of matrix and its related method. Also it has many other desirable error propagation properties. For these reason, it has become one of the popular orthogonal transformations and a standard tool of numerical linear algebra. An important application of Householder reflectors is found in the implementation of the implicit QR algorithm for the solution of the Hessenberg eigenvalue problem [1],[6].

Householder transformation has the canonical form

$$P = I - \frac{2}{v^*v}vv^* \qquad where\ v \in C^n. \tag{1.1}$$

Received by the editors on December 7, 2001.
Key words and phrases: Householder, QR factorization.

From computational viewpoint, Householder transformations are used to selectively zero out blocks of entries in vectors or in columns of matrices and are represented by the isometric mapping of a driving vector z into a stretching of a vector of the canonical basis

$$P_k z = \sigma e_k, \qquad |\sigma| = \| z \|,$$

where $z \in C^n$ and e_k is kth column of identity matrix in $C^{n\times n}$. For this purpose, the transformation matrix P_k can be represented by

$$P_k = I - \frac{(z - \sigma e_k)(z - \sigma e_k)^*}{\bar{\sigma}(\sigma - z_k)} \tag{1.2}$$

or

$$P_k = I - uu^*, \qquad u = \frac{z - \sigma e_k}{[\bar{\sigma}(\sigma - z_k)]^{1/2}}, \quad \| u \| = \sqrt{2} \tag{1.3}$$

depending on the computational advantage of accurate square roots. We define, here, σ with its modulus and the hermiticity relation

$$\bar{\sigma} z_k = \pm | z_k | \| z \|. \tag{1.4}$$

As shown in section 2, we have two choices of σ to make Householder transformation from (1.4). Dubrulle provides an analysis in [3] that the one type of two choices of σ, which is known as unstable type, has better ability to propagate the information borne by its driving vector. Two types of Householder matrices are defined and some extendable known results are summarized in section 2.

In this paper, a new analysis of QR factorizations using the two types of Householder reflections Parllett defined is presented. It shows that the use of the common type(the first) in QR factorizations with pivoting for size by row and column exchanges leads to rowwise more stability, although the second type has been used in Householder

transformation for the better purpose. A row-oriented backward error bound for the second type Householder QR factorizations with column pivoting is drived in section 3.

We developed the Matlab code for Householder transformation and QR factorization to test our error analysis for some examples. The computational results are shown in section 4.

2. Properties for two types of Householder Transformation

From the formula (1.4) we obtain two types of σ,

$$\sigma = \pm \frac{z_k}{|z_k|}\|z\|, \qquad z_k \neq 0$$

with the arbitrary choice $\sigma = \|z\|$ when $z_k = 0$. In general, we have been to use the minus sign because of the stability of Householder transformation, the formula (1.2). Parlett, however, showed that decision of stability of (1.2) is depending on the calculation of $(\sigma - z_k)$ rather than the choice of sign. From the stable computation of $(\sigma - z_k)$, the two types can be stable in norm. We summarize the different roles of z_k and $\{z_i\}_{i \neq k}$ in the two types of transformations. And the properties that reflectors carry information of z.

In the case of $\sigma = -\frac{z_k}{|z_k|}\|z\|$, from the formula (1.2) and (1.3), we obtain

$$P_k = I - \frac{(z - \sigma e_k)(z - \sigma e_k)^*}{\|z\|(\|z\| + |z_k|)} = I - uu^*. \tag{2.1}$$

Here, vector u is composed of different element,

$$u_i = \begin{cases} \dfrac{z_k}{|z_k|}\left(1 + \dfrac{|z_k|}{\|z\|}\right)^{1/2} & i = k \qquad (2.2.1) \\ \dfrac{z_i}{\|z\|}\left(1 + \dfrac{|z_k|}{\|z\|}\right)^{-1/2} & i \neq k \qquad (2.2.2) \end{cases}$$

For the case of $\sigma = \frac{z_k}{|z_k|}\|z\|$, from the formula (1.2) and (1.3), and by the stable computation of $\sigma - z_k = \frac{\|z - z_k e_k\|^2 z_k}{(\|z\| + |z_k|)|z_k|}$, we have

$$P_k = I - \frac{1 + \frac{|z_k|}{\|z\|}}{\|z - z_k e_k\|^2}(z - \sigma e_k)(z - \sigma e_k)^* = I - uu^*, \qquad (2.3)$$

$$u_i = \begin{cases} -\frac{z_k}{|z_k|}\frac{\|z - z_k e_k\|}{\|z\|}(1 + \frac{|z_k|}{\|z\|})^{-1/2} & i = k \qquad (2.4.1) \\ \frac{z_i}{\|z - z_k e_k\|}(1 + \frac{|z_k|}{\|z\|})^{1/2} & i \neq k. \qquad (2.4.2) \end{cases}$$

The first type shows that z_k is relatively more important than z_k in the second type, specially for $i \neq k$. From the propagating view of the information of vector z, Dubrulle [3] examined how well elementary reflectors carry information using the formulas (2.2) and (2.4).

Consider the reflector in form (1.3) defined by

$$P_k = I - uu^*, \quad P_k z = \sigma e_k, \quad |\sigma| = \| z \|, \quad y = P_k x = x_i - (u^* x) u_i,$$

for some nonzero arbitrary vector x. The following conditions for which the floating-point invariance of $x_i, i \neq k$ under P_k occurs can be derived for the two types of elementary reflectors :

$$\begin{cases} \frac{|z_i|}{\|z\|} \leqslant \frac{\epsilon}{\sqrt{2}}\sqrt{1 + \frac{|z_k|}{\|z\|}}\frac{|x_i|}{\|x\|} & if\ \sigma = -\frac{z_k}{|z_k|}\|z\| \quad (2.5) \\ \frac{|z_i|}{\|z - z_k e_k\|}\sqrt{1 + \frac{|z_k|}{\|z\|}} \leqslant \frac{\epsilon}{\sqrt{2}}\frac{|x_i|}{\|x\|} & if\ \sigma = \frac{z_k}{|z_k|}\|z\|, \quad (2.6) \end{cases}$$

where $\frac{|y_i - x_i|}{|x_i|} \leqslant \epsilon$, $x_i \neq 0$, $i \neq k$. From the above inequalities, if $\|z_k\| \approx \|z\|$, than $\|z - z_k e_k\| \to 0$ and hence the formula (2.6) is

harder to satisfy then the formula (2.5). It means the second type is better at propagating the information contained in each z_i for $i \neq k$.

3. Stability for two types of Householder QR factorization.

QR factorization provides a standard way to solve the least squares problem, and the QR factorization is perhaps most often computed using Householder transformation (as is done in LINPACK and LAPACK, for example). This secition derives a row-oriented backward error bound for Householder QR factorization with column pivoting based of the description of two types in section 2. We examine the error bound of Householder QR factorization for the first type is still better than one for the second type although the second type is better in the propagation of information about z.

Let $A = A^{(1)} \in C^{m\times n}(m \geqslant n)$ and let $a_j^{(k)}$ denote the jth column of $A^{(k)}$, the reduced matrix at the start of the kth stage of the reduction to trapezoidal form. Also we form the Householder matrix from the formula (1.1) and (1.2)

$$P_k = I - \beta_k v_k v_k^* \in C^{m\times n}, \quad \beta_k = \frac{2}{v_k^* v_k}, \tag{3.1}$$

where $v_k(1 : k-1) = 0$ and

$$v_k(k : m) = a_k^{(k)}(k : m) - \sigma_k e_1, \tag{3.2}$$

where $e_1 \in R^{m-k+1}$ is the first unit vector and

$$\sigma_k = \pm \frac{a_{kk}^{(k)}}{|\, a_{kk}^{(k)}|} \|a_k^{(k)}(k : m)\|_2, \qquad a_{kk}^{(k)} \neq 0.$$

The Householder matrix P_k has the property that $a_k^{(k+1)} = P_k a_k^{(k)}$ satisfies $a_k^{(k+1)}(k : m) = \sigma_k e_1$. In QR factorization with column pivoting, columns are exchanged at the start of the kth stage to ensure

that

$$|\sigma_k| = \|a_k^{(k)}(k:m)\|_2 = \max_{j \geqslant k} \|a_j^{(k)}(k:m)\|_2, \tag{3.3}$$

$$|\sigma_1| \geqslant |\sigma_2| \geqslant \cdots \geqslant |\sigma_n|.$$

To simplify the notation we assume, without loss of generality, that A is prepivoted, that is, that no column interchanges are required in order to satisfy (3.3). The error bound of QR factorization for the first type is well described in [2] although the analysis is not based on the in section 2 results. In the rest of section we examine the error bound for the second type which is not described in [2].

The following lemma will be used to analyze the error for the second type, i.e. $\sigma_k = \frac{a_{kk}^{(k)}}{|a_{kk}^{(k)}|} \| a_k^{(k)}(k:m)\|_2$.

LEMMA 3.1. *If* $\sigma_k = \frac{a_{kk}^{(k)}}{|a_{kk}^{(k)}|} \| a_k^{(k)}(k:m)\|_2$ *and* $A^{(k+1)} = P_k A^{(k)}$ *then* $a_j^{(k+1)} = a_j^{(k)} - \phi_j^{(k)} v_k$, $j \geqslant k$ *where* $\phi_j^{(k)} = \beta_k v_k^* a_j^{(k)}$ *satisfies*

$$|\phi_j^{(k)}| \leqslant 2 \frac{\| a_k^{(k)}(k:m)\|_2}{\| a_k^{(k)}(k+1:m)\|_2}. \tag{3.4}$$

Proof. By the formula (1.2) and (2.3),

$$\begin{aligned} \frac{\|v_k\|_2^2}{2} &= \frac{\| a_k^{(k)}(k+1:m)\|_2^2}{1 + \frac{|a_{kk}^{(k)}|}{\| a_k^{(k)}(k:m)\|_2}} \\ &= \frac{|\sigma_k| \| a_k^{(k)}(k+1:m)\|_2^2}{|\sigma_k| + |a_{kk}^{(k)}|} \end{aligned} \tag{3.5}$$

Also, since $|\sigma_k| = \| a_k^{(k)}(k:m)\|_2 = \max_{j \geqslant k} \| a_j^{(k)}(k:m)\|_2$, and

$\phi_j^{(k)} = \beta_k v_k^* a_j^{(k)} = \beta_k v_k (k:m)^* a_j^{(k)}(k:m)$, $\phi_j^{(k)}$ satisfies

$$|\phi_j^{(k)}| \leqslant |\beta_k| \|v_k\|_2 \|a_j^{(k)}(k:m)\|_2 \tag{3.6}$$

$$= \frac{2\|a_j^{(k)}(k:m)\|_2}{\|v_k\|_2} = \frac{2\| a_k^{(k)}(k:m)\|_2}{(\frac{2|\sigma_k|\| a_k^{(k)}(k+1:m)\|_2^2}{|\sigma_k|+|a_{kk}^{(k)}|})^{1/2}}$$

$$\leqslant \sqrt{2}(1+\frac{|a_{kk}^{(k)}|}{|\sigma_k|})^{1/2}(\frac{\| a_j^{(k)}(k:m)\|_2}{\| a_k^{(k)}(k+1:m)\|_2})$$

$$\leqslant 2\frac{\| a_k^{(k)}(k:m)\|_2}{\| a_k^{(k)}(k+1:m)\|_2}.$$

□

The problem can occur whenever the leading column $a_k^{(k)}$ is such that $\|a_{kk}^{(k)}\| \approx \|a_k^{(k)}\|$. Hence $|\phi_j^{(k)}| \leqslant 2|a_{kk}^{(k)}|$ can be large.

THEOREM 3.2. *Let $\hat{R} \in C^{m\times n}$ be the computed upper trapezoidal QR factor of $A \in C^{m\times n}(m \geqslant n)$ obtained via the Householder QR Algorithm with column pivoting. Then there exists an orthogonal $Q \in C^{m\times m}$ such that*

$$(A + \Delta A)\Pi = Q\hat{R},$$

where Π is a permutation matrix that describes the overall effect of the column interchanges and

$$|\Delta A| \leqslant \tilde{r}_m \rho^2 \Lambda e^\top D, \tag{3.7}$$

where $\Lambda = (\lambda_1\ \lambda_2\ \cdots\ \lambda_m)^\top$, $\lambda_i = \max_{j,k}|\hat{a}_{ij}^{(k)}|$, $e = (1\ 1\cdots 1)^\top$, $D = diag(1^2, 2^2, \cdots, n^2)$, and $\rho = \frac{\| a_k^{(k)}(k:m)\|_2}{\| a_k^{(k)}(k+1:m)\|_2}$.

Proof. From the proof of lemma 3.1 and (3.4), we have $|\beta_k|\,|v_k|^*\,|\hat{a}_j^{(k)}| \leqslant 2\rho$. Using standard error result,

$$\hat{a}_j^{(k+1)} = (P_k + \Delta P_k)\hat{a}_j^{(k)} = P_k \hat{a}_j^{(k)} + f_j^{(k)}, \tag{3.8}$$

where $f_j^{(k)} = \Delta P_k \hat{a}_j^{(k)}$. Also, since $\hat{a}_j^{(k+1)}(1:k-1) = \hat{a}_j^{(k)}(1:k-1)$, $f_j^{(k)}(1:k-1) = 0$ and

$$\begin{aligned} |f_j^{(k)}| &\leqslant u|\hat{a}_j^{(k)}| + \tilde{r}_{m-k}(|\beta_k|\,|v_k|^*\,|\hat{a}_j^{(k)}|)|v_k| \\ &\leqslant u|\hat{a}_j^{(k)}| + \rho\tilde{r}_{m-k}|v_k|, \end{aligned} \tag{3.9}$$

where u is the unit roundoff. Let $\lambda_i = \max_{j,k} |\hat{a_{ij}}^{(k)}|$, $\Lambda = (\lambda_1 \ \lambda_2 \cdots \lambda_m)^\top$. Then by $|\hat{a}_j^{(k)}| \leqslant \Lambda$ and

$$|v_k|_i \leqslant \left\{ \begin{array}{ll} \lambda_k + |\sigma_k| \leqslant 2\lambda_k & i = k, \\ \lambda_i & i > k \end{array} \right\} \leqslant 2\Lambda \tag{3.10}$$

we obtain $|f_j^{(k)}| \leqslant u\Lambda + 2\rho\tilde{r}_{m-k}\Lambda = (u + \rho\tilde{r}_{m-k})\Lambda$. And by (3.5) we have

$$\|v_k\|_2 \geqslant \|a_k^{(k)}(k+1:m)\|_2. \tag{3.11}$$

Now from (3.9) using (3.11) and (3.2)

$$\begin{aligned} \frac{\|f_j^{(i)}\|_2}{\|v_k\|_2} &\leqslant u\frac{\|\hat{a}_j^{(i)}(i:m)\|_2}{\|v_k\|_2} + \rho\tilde{r}_{m-i}\frac{\|v_i\|_2}{\|v_k\|_2} \\ &\leqslant u\frac{|\sigma_i|}{\|a_k^{(k)}(k+1:m)\|_2} + \rho\tilde{r}_{m-i}\frac{|\sigma_i|}{\|a_k^{(k)}(k+1:m)\|_2} \\ &\leqslant u\rho + \rho^2\tilde{r}_{m-i} = \rho^2\tilde{r}_{m-i}. \end{aligned}$$

Also by error analysis for the QR factorization and (3.8)

$$\hat{a}_j^{(k)} = P_k\hat{a}_j^{(k+1)} - P_k f_j^{(k)}.$$

This is presented the following formula

$$\begin{aligned} \hat{a}_j^{(1)} &= P_1\hat{a}_j^{(2)} - P_1 f_j^{(1)} = P_1(P_2\hat{a}_j^{(3)} - P_2 f_j^{(2)}) - P_1 f_j^{(1)} \\ &= \cdots = P_1P_2\cdots P_j\hat{a}_j^{(j+1)} - P_1P_2\cdots P_j f_j^{(j)} - \cdots - P_1 f_j^{(1)}. \end{aligned}$$

Since $a_j = \hat{a}_j^{(1)}$ and $\hat{a}_j^{(j+1)} = \hat{a}_j^{(n+1)}$,

$$a_j = P_1 P_2 \cdots P_j \, \hat{a}_j^{(n+1)} - \sum_{i=1}^{j} P_1 P_2 \cdots P_i f_j^{(i)} \tag{3.12}$$

Now, let $y_i = P_1 P_2 \cdots P_i f_j^{(i)}, i \leqslant j$, then we have

$$y_i = (I - \beta_1 v_1 v_1^*) P_2 P_3 \cdots P_i f_j^{(i)} = P_2 P_3 \cdots P_i f_j^{(j)} - \beta_1 v_1 v_1^* P_2 P_3 \cdots P_i f_j^{(i)}$$
$$= \cdots = f_j^{(i)} - \sum_{k=1}^{i} \beta_k v_k v_k^* P_{k+1} P_{k+2} \cdots P_i f_j^{(i)}.$$

Writing $z_k = \beta_k v_k v_k^* P_{k+1} P_{k+2} \cdots P_i f_j^{(i)} = \frac{2 v_k v_k^*}{v_k^* v_k} P_{k+1} P_{k+2} \cdots P_i f_j^{(i)}$ and using (3.10), we have $|z_k| \leqslant 4\Lambda \frac{\|f_j^{(i)}\|_2}{\|v_k\|_2} \quad k \leqslant i$. Thus we conclude that

$$|y_i| \leqslant (u + \rho \tilde{r}_{m-i})\Lambda + 4i\rho^2 \tilde{r}_{m-i}\Lambda = i\rho^2 \tilde{r}_{m-i}\Lambda.$$

And from (3.12),

$$a_j = P_1 P_2 \cdots P_j \, \hat{a}_j^{(n+1)} + h_j,$$

where $|h_j| \leqslant \sum_{i=1}^{j} i\rho^2 \tilde{r}_{m-i}\Lambda = j^2 \rho^2 \tilde{r}_m \Lambda$. Also,

$$|\Delta A| = |Q\hat{R} - A\Pi| = [\, |h_1| \;\; |h_2| \;\; \cdots \;\; |h_n| \,]$$
$$\leqslant \rho^2 \tilde{r}_m \Lambda e^\top D$$

□

We can rewrite (3.7) in the slightly weakened form

$$\max_i \frac{\|\Delta A(i,:)\|_\infty}{\|A(i,:)\|_\infty} \leqslant \max_i \frac{n^2 \rho^2 \tilde{r}_m \alpha_i}{\|A(i,:)\|_\infty}. \tag{3.13}$$

Thus, the use of second type in QR factorizations may lead to instability whenever the leading column $a_k^{(k)}$ is such that $\|a_{kk}^{(k)}\| \approx \|a_k^{(k)}\|$.

4. Numerical Examples and Experimental Results.

We implemented the Householder transformation and QR factorization with Matlab and ran them on the sun workstation to check the error for the both types described in the previous sections.

The following example [3], where the Householder vectors are computed with the Matlab function algorithm, illustrates conclusion of the section 2.

EXAMPLE 4.1 Let $z = [1 \ 6\eta \ 2\eta]^\top$, $\eta = \frac{\epsilon}{8}$, $\epsilon = 10^{-8}$, $x = [1 \ 1 \ 1]^\top$, $k = 1$ Using the first type, the invariance criterion is satisfied for $i > 1$, and the information borne by z_2 and z_3 is lost in the floating-point representation $fl(y) = [-1 \ 1 \ 1]^\top$ of y, while the invariance test for the second type is not satisfied, and the transformation preserves all subspace information $fl(y) = [1 + \epsilon \ -1.4 \ 0.2]^\top$ This phenomenon shows that the second type is better at propagating the information contained in each z_i for $i \neq k$.

But the following example shows that the use of the second type in QR factorization may lead to rowwise instability.

EXAMPLE 4.2

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 6\eta & 1 & \eta \\ 2\eta & 1 & 1 \\ 3\eta & \eta & -2\eta \end{pmatrix} \qquad \eta = \frac{\epsilon}{8}, \quad \epsilon = 10^{-8}$$

Testing with Matlab code, the rowwise error of Householder QR factorization was 9.2830×10^{-16} for the first type, but 4.7696×10^{-8} for the second type.

This example implies that the second type in QR factorizations may lead to rowwise instability, although the second type is propagating better information of z.

5. Conclusion.

The analysis of Householder transformations in section 2 reveals that the second type has a better ability to propagate the information borne by its driving vector. However, the use of the second type in QR factorizations with pivoting for size by row and column exchanges may lead to rowwise instability, although the Householder algorithm is stable. We suggest that the QR factorization should be used by the first type of transformation even though the second type has been used in the Householder transformation for the any purpose.

References

1. Z.Bai and J.Demmel, *On a Block Implemention of Hessenberg multishift QR Iteration*, Int. J. High-Speed Comput. **62** (1989), 209-226.
2. A.Cox and N.Higham, *Stability of Householder QR Factorization for Weighted Least-Squares Problem*, Numerical Analysis TR 301, Dept. of Mathematics, University of Manchester,UK (1997).
3. A.A.Dubrulle, *Householder Transformations Revisited*, SIAM J. Matrix Anal. Appl Math. **22(1)** (2000), 33-40.
4. C.Davis and W.Kahan, *Some New Bounds on Perturbation of Subspaces*, BULL. Amer. Math. Soc. **75** (1969), 863-868.
5. A.A.Dubrulle, *A QR Algorithm with Variable Iteration Multiplicity*, J. Comp. Appl. Math. **86** (1997), 125-139.
6. A.A.Dubrulle, *The Multishift QR Algorithm : Is It Worth the Trouble?*, TR G320-3588, IBM Scientific Center, Palo Alot, CA, (1991).
7. B.Danloy, *On the Choice of Signs for Householder's Matrices*, J. Comp. Appl. Math. **2 (1)** (1976), 67-69.
8. G.H.Golub and C.F.Van.Loan, *Matrix Computation*, The Johns Hopkins University Press.
9. Nicholas J.Higham, *Accuracy and Stability of Numerical Algorithm..*
10. B.N.Parlett, *Analysis of Algorithm for Reflection in Bisector*, SIAM Rev., **13** (1971), 197-208.

*

Department of Mathematics
Chungnam National University
Taejon 305-764, Korea
E-mail: yja@math.cnu.ac.kr

**
Department of Mathematics
Chungnam National University
Taejon 305-764, Korea
E-mail: soh@math.cnu.ac.kr

***
Department of Mathematics
Chungnam National University
Taejon 305-764, Korea
E-mail: sychung@math.cnu.ac.kr