A Perturbation Analysis For R in the QR Factorization

Xiao-Wen Chang   *Christopher C. Paige*

Technical Report No. SOCS-95.7
November 1995

A PERTURBATION ANALYSIS FOR R IN THE QR FACTORIZATION *

XIAO-WEN CHANG† AND CHRISTOPHER C. PAIGE†

Abstract. We present new normwise and componentwise perturbation analyses for the R factor of the QR factorization $A = Q_1R$ of an $m \times n$ matrix A with full column rank. The analyses more accurately reflect the sensitivity of the problem than previous normwise and componentwise results. The new condition numbers here are altered by any column pivoting used in $AP = Q_1R$, and are bounded for a fixed n when the standard column pivoting strategy is used. Both numerical results and an analysis show that the standard method of pivoting is optimal in that it usually leads to a normwise condition number very close to its lower limit for any given A. It follows that the computed R will probably have greatest accuracy when we use the standard column pivoting strategy. Also we derive a practical estimate for the normwise condition number.

Key words. QR factorization, perturbation analysis, pivoting

AMS Subject Classifications: 65F05, 65F30, 65G05

1. Introduction.

The QR factorization is an important tool in matrix computations (see [2]): given an $m \times n$ real matrix A with full column rank, there exists a unique $m \times n$ matrix Q_1 with orthonormal columns, and a unique nonsingular upper triangular R with positive diagonal entries such that

$$A = Q_1R.$$

The matrix Q_1 is referred to as the orthogonal factor, and R the triangular factor.

The perturbation analysis for the QR factorization has been considered by several authors. The first norm-based result for R was presented by Stewart [5]. That was further modified and improved by Sun [10]. Using different approaches Sun [10] and Stewart [6] gave *first order* normwise perturbation analyses for R. First order and strict componentwise perturbation analyses for R have been given by Zha [13] and Sun [11], respectively.

A derivation of this first order normwise perturbation result for R follows.

THEOREM 1.1. [10]. *Let $A \in \mathcal{R}^{m\times n}$ be of full column rank, with the QR factorization $A = Q_1R$, and let ΔA be a real $m \times n$ matrix. If $\epsilon = \|\Delta A\|_F/\|A\|_2$ satisfies*

$$\kappa_2(A)\epsilon \le 1, \tag{1.1}$$

then $A + \Delta A$ has a unique QR factorization

$$A + \Delta A = (Q_1 + \Delta Q_1)(R + \Delta R),$$

where

$$\frac{\|\Delta R\|_F}{\|R\|_2} \le \sqrt{2}\kappa_2(A)\epsilon + O(\epsilon^2), \tag{1.2}$$

with $\kappa_2(A) \equiv \|A\|_2\|A^\dagger\|_2$.

---

* This research was partially supported by NSERC of Canada Grant OGP0009236.

† School of Computer Science, McGill University, Montreal, Quebec, Canada, H3A 2A7, (chang@cs.mcgill.ca), (chris@cs.mcgill.ca).

Proof. Let $G \equiv \Delta A/\epsilon$ (if $\epsilon = 0$ the theorem is trivial). If (1.1) holds, then $A + tG$ has full column rank for all $t \in [0, \epsilon]$, and so has the unique QR factorization

$$A(t) \equiv A + tG = Q_1(t)R(t), \tag{1.3}$$

where

$$Q_1^T(t)Q_1(t) = I. \tag{1.4}$$

Notice that $R(0) = R$ and $R(\epsilon) = R + \Delta R$.

It can easily be verified from the algorithm for the QR factorization that $Q_1(t)$ and $R(t)$ are twice continuously differentiable for $t \in [0, \epsilon]$. Using (1.3) and (1.4) we have

$$A(t)^T A(t) = R(t)^T R(t).$$

Differentiating this and the first equality of (1.3), and setting $t = 0$, gives

$$R^T \dot{R}(0) + \dot{R}^T(0)R = A^T \dot{A}(0) + \dot{A}(0)^T A, \qquad \dot{A}(0) = G.$$

Combining these with $A = Q_1 R$ gives the key equation for $\dot{R}(0)$

$$R^T \dot{R}(0) + \dot{R}^T(0)R = R^T Q_1^T G + G^T Q_1 R, \tag{1.5}$$

$$\dot{R}(0)R^{-1} + (\dot{R}(0)R^{-1})^T = Q_1^T G R^{-1} + (Q_1^T G R^{-1})^T. \tag{1.6}$$

It follows that $(\dot{R}(0)R^{-1})_{ii} = (Q_1^T G R^{-1})_{ii}$, and since $\dot{R}(0)R^{-1}$ is upper triangular, $(\dot{R}(0)R^{-1})_{ij} = (Q_1^T G R^{-1})_{ij} + (Q_1^T G R^{-1})_{ji}$ for $i < j$, giving

$$\|\dot{R}(0)R^{-1}\|_F^2 \le 2\|Q_1^T G R^{-1}\|_F^2, \tag{1.7}$$

$$\|\dot{R}(0)R^{-1}\|_F \le \sqrt{2}\|R^{-1}\|_2 \|Q_1^T G\|_F \le \sqrt{2}\|R^{-1}\|_2 \|G\|_F. \tag{1.8}$$

Notice that $\|R\|_2^{-1}\|\dot{R}(0)\|_F \le \|\dot{R}(0)R^{-1}\|_F$, so

$$\|R\|_2^{-1}\|\dot{R}(0)\|_F \le \sqrt{2}\|R^{-1}\|_2 \|G\|_F,$$

which, together with $\|G\|_F = \|A\|_2 = \|R\|_2$ and $\|R^{-1}\|_2 = \|A^\dagger\|_2$, gives

$$\frac{\|\dot{R}(0)\|_F}{\|R\|_2} \le \sqrt{2}\kappa_2(A). \tag{1.9}$$

The Taylor expansion for $R(t)$ about $t = 0$ gives at $t = \epsilon$

$$R + \Delta R = R(\epsilon) = R(0) + \epsilon \dot{R}(0) + O(\epsilon^2), \tag{1.10}$$

so that

$$\frac{\|\Delta R\|_F}{\|R\|_2} \le \frac{\|\dot{R}(0)\|_F}{\|R\|_2}\epsilon + O(\epsilon^2),$$

which, combined with (1.9), gives (1.2). $\square$.

The proof of Theorem 1.1 shows that the key point in the derivation of a first order perturbation bound is the use of (1.5) to give a good bound on the sensitivity $\|\dot{R}(0)\|_F/\|R\|_2$.

The first order componentwise perturbation analysis for R can be summarized as follows.

THEOREM 1.2. [13]. *Let $A \in \mathcal{R}^{m\times n}$ be of full column rank n with the QR factorization $A = Q_1 R$, and let ΔA with $|\Delta A| \leq \epsilon C|A|$ be such that $A+\Delta A$ is also of full column rank, where C is nonnegative with $c_{ij} \leq 1$. Denote by $\|\cdot\|$ any consistent monotone matrix norm, that is $|A| \leq |B|$ implies $\|A\| \leq \|B\|$, and $\|AB\| \leq \|A\|\|B\|$ (see for example* [9, p. 52]*). Let*

$$\kappa_{BS}(S) \equiv \||S^{-1}|\cdot|S|\| \tag{1.11}$$

be the Bauer-Skeel condition number, and

$$\eta = \max(\||Q^T|C|Q|\|, \||Q^T|C^T|Q|\|).$$

If

$$\epsilon\eta[\kappa_{BS}(R^{-1}) + \kappa_{BS}(R^T)] < 1,$$

then

$$\frac{\|\Delta R\|}{\|R\|} \leq \epsilon\eta[\kappa_{BS}(R^{-1}) + \kappa_{BS}(R^T)] + O(\epsilon^2). \tag{1.12}$$

Zha suggested using $\kappa_S(A) \equiv (\kappa_{BS}(R^{-1}) + \kappa_{BS}(R^T))/2$ as a condition number for the QR factorization under the columnwise class of perturbation considered. Notice $\kappa_S(A)$ is independent of the column scaling of A.

The purpose of this paper is to establish new first order perturbation bounds for the R factor, which are generally sharper than the equivalent results of [10, 13].

2. New perturbation bounds. For any matrix $C \equiv (c_{ij}) \equiv [c_1, \ldots, c_n] \in \mathcal{R}^{n\times n}$, denote by $c_j^{(i)}$ the vector of the first i elements of c_j. With this, we define (using "u" to denote "upper")

$$\mathrm{uvec}(C) \equiv \begin{bmatrix} c_1^{(1)} \\ c_2^{(2)} \\ \cdot \\ c_n^{(n)} \end{bmatrix}.$$

It is the vector formed by stacking the columns of the upper triangular part of C into one long vector.

In order to estimate $\dot{R}(0)$ using (1.5), as in Stewart [5] we define the linear operator $\mathbf{T}_R$ that maps the space of upper triangular matrices into the space of symmetric matrices by

$$\mathbf{T}_R X = R^T X + X^T R,$$

where $R \in \mathcal{R}^{n\times n}$ is nonsingular upper triangular. We also define the operator $\hat{\mathbf{T}}_R$, an extension of $\mathbf{T}_R$, which maps the space of real square matrices into the space of symmetric matrices by

$$\hat{\mathbf{T}}_R F = R^T F + F^T R.$$

To analyse (1.5) we consider the equation $\mathbf{T}_R X = \hat{\mathbf{T}}_R F$, i.e.,

$$R^T X + X^T R = R^T F + F^T R. \tag{2.1}$$

The upper and lower triangular parts of (2.1) contain identical information, and we now write the upper triangular part of (2.1) in matrix–vector, rather than matrix–matrix format. The first j elements of the jth column of this is

$$R_{jj}^T x_j^{(j)} + X_{jj}^T r_j^{(j)} = R_{jj}^T f_j^{(j)} + \begin{bmatrix} f_1^{(j)T} \\ f_2^{(j)T} \\ \cdot \\ f_j^{(j)T} \end{bmatrix} r_j^{(j)},$$

and by rewriting this, we can see how to solve for $x_j^{(j)}$, $j = 1, \ldots, n$:

$$(R_{jj}^T + e_j r_j^{(j)T}) x_j^{(j)} + \begin{bmatrix} r_j^{(1)T} x_1^{(1)} \\ r_j^{(2)T} x_2^{(2)} \\ \cdot \\ r_j^{(j-1)T} x_{j-1}^{(j-1)} \\ 0 \end{bmatrix} = (R_{jj}^T + e_j r_j^{(j)T}) f_j^{(j)} + \begin{bmatrix} r_j^{(j)T} f_1^{(j)} \\ r_j^{(j)T} f_2^{(j)} \\ \cdot \\ r_j^{(j)T} f_{j-1}^{(j)} \\ 0 \end{bmatrix},$$

which, on dividing the last row of this by 2, gives

$$W_R \mathrm{uvec}(X) = Z_R \mathrm{vec}(F) \tag{2.2}$$

where $W_R \in \mathcal{R}^{\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}}$ is

$$\left[\begin{array}{c|cc|ccc|c|ccccc}
r_{11} & & & & & & & & & & & \\
r_{12} & r_{11} & & & & & & & & & & \\
 & r_{12} & r_{22} & & & & & & & & & \\
r_{13} & & & r_{11} & & & & & & & & \\
 & r_{13} & r_{23} & r_{12} & r_{22} & & & & & & & \\
 & & & r_{13} & r_{23} & r_{33} & & & & & & \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & & & & \\
r_{1n} & & & & & & & r_{11} & & & & \\
 & r_{1n} & r_{2n} & & & & & r_{12} & r_{22} & & & \\
 & & & r_{1n} & r_{2n} & r_{3n} & & r_{13} & r_{23} & r_{33} & & \\
 & & & & & & \cdot & \cdot & \cdot & \cdot & \cdot & \\
 & & & & & & & r_{1n} & r_{2n} & r_{3n} & \cdot & r_{nn}
\end{array}\right],$$

and $Z_R \in \mathcal{R}^{\frac{n(n+1)}{2} \times n^2}$ is

$$\left[\begin{array}{cccc|cccc|c|cccc}
r_{11} & & & & & & & & & & & & \\
\hline
r_{12} & r_{22} & & & r_{11} & & & & & & & & \\
 & & & & r_{12} & r_{22} & & & & & & & \\
\hline
\cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot & & & & \\
\hline
r_{1n} & r_{2n} & \cdot & r_{nn} & & & & & & r_{11} & & & \\
 & & & & r_{1n} & r_{2n} & \cdot & r_{nn} & & r_{12} & r_{22} & & \\
 & & & & & & & & \cdot & \cdot & \cdot & \cdot & \\
 & & & & & & & & & r_{1n} & r_{2n} & \cdot & r_{nn}
\end{array}\right].$$

Since R is nonsingular, W_R is also, and from (2.2)

$$\text{uvec}(X) = W_R^{-1} Z_R \text{vec}(F). \tag{2.3}$$

so that for any matrix F, $\mathbf{T}_R X = \hat{\mathbf{T}}_R F$ has a unique (upper triangular) solution, and the operator $\mathbf{T}_R$ is nonsingular.

By comparing (1.5) with (2.1) and its equivalent (2.3) we see that $W_R^{-1} Z_R$ is crucial to our analysis. Thus for our norm-based perturbation analysis we want to know something about $\|W_R^{-1} Z_R\|_2$. To do this we first define the operator norm subordinate to the matrix Frobenius norm

$$\|\mathbf{T}_R^{-1}\hat{\mathbf{T}}_R\|_F \equiv \sup_{F \in \mathcal{R}^{n\times n}} \frac{\|\mathbf{T}_R^{-1}\hat{\mathbf{T}}_R F\|_F}{\|F\|_F}.$$

It is easy to show

$$\|\mathbf{T}_R^{-1}\hat{\mathbf{T}}_R\|_F = \|W_R^{-1} Z_R\|_2. \tag{2.4}$$

With any matrix F, $\mathbf{T}_R X = \hat{\mathbf{T}}_R F$ gives $XR^{-1} + (XR^{-1})^T = FR^{-1} + (FR^{-1})^T$, and following the same argument for (1.8), $\|XR^{-1}\|_F \le \sqrt{2}\|R^{-1}\|_2\|F\|_F$, so that

$$\|\mathbf{T}_R^{-1}\hat{\mathbf{T}}_R F\|_F = \|X\|_F = \|XR^{-1}R\|_F \le \|XR^{-1}\|_F\|R\|_2 \le \sqrt{2}\kappa_2(R)\|F\|_F. \tag{2.5}$$

Combining this with (2.4), we get

$$\|W_R^{-1} Z_R\|_2 = \sup_{F \in \mathcal{R}^{n\times n}} \frac{\|\mathbf{T}_R^{-1}\hat{\mathbf{T}}_R F\|_F}{\|F\|_F} \le \sqrt{2}\kappa_2(R). \tag{2.6}$$

Note that the above upper bound is tight, in the sense that the equality will hold if R is an $n \times n$ identity matrix with $n \ge 2$. It is easy to observe that each column of W_R is one of columns of Z_R. Thus $W_R^{-1} Z$ has an $n(n+1)/2 \times n(n+1)/2$ identity submatrix. It follows that

$$\|W_R^{-1} Z_R\|_2 \ge 1. \tag{2.7}$$

This lower bound is approximately tight for any n, in the sense that by taking $R = \text{diag}(1, \epsilon, \ldots, \epsilon^{n-1})$, $\|W_R^{-1} Z_R\|_2 \to 1$ as $\epsilon \to 0$.

This analysis leads to new first order, normwise and componentwise perturbation bound results.

THEOREM 2.1. *Let $A = Q_1 R$ be the QR factorization of $A \in \mathcal{R}^{m\times n}$ with full column rank, and let ΔA be a real $m \times n$ matrix. If $\epsilon = \|\Delta A\|_F / \|A\|_2$ satisfies $\kappa_2(A)\epsilon \le 1$, then there is a unique QR factorization*

$$A + \Delta A = (Q_1 + \Delta Q_1)(R + \Delta R), \tag{2.8}$$

where

$$\frac{\|\Delta R\|_F}{\|R\|_2} \le \kappa_R(A)\epsilon + O(\epsilon^2), \tag{2.9}$$

and

$$1 \le \kappa_R(A) \equiv \|W_R^{-1} Z_R\|_2 \le \sqrt{2}\kappa_2(A). \tag{2.10}$$

Proof. By the same argument used in the proof of Theorem 1.1, $A + \Delta A$ has the QR factorization (2.8). From (1.5), (2.1), (2.3) and the expression for G we have

$$\mathrm{uvec}(\dot{R}(0)) = W_R^{-1} Z_R \mathrm{vec}(Q_1^T \Delta A/\epsilon),$$

so taking the 2-norm gives

$$\|\dot{R}(0)\|_F \leq \|W_R^{-1} Z_R\|_2 \|Q_1^T \Delta A/\epsilon\|_F \leq \|W_R^{-1} Z_R\|_2 \|A\|_2.$$

Combining this with $\|A\|_2 = \|R\|_2$, $\|A^\dagger\|_2 = \|R^{-1}\|_2$, (2.6) and (2.7) we get (2.10) and

$$\frac{\|\dot{R}(0)\|_F}{\|R\|_2} \leq \kappa_R(A). \tag{2.11}$$

Thus, from the Taylor series (1.10) of $R(t)$, (2.9) follows. □

Remark 1. From (2.10) we know the first order perturbation bound (2.9) is at least as sharp as (1.2), but it suggests it may be considerably sharper. In fact it can be sharper by an arbitrary factor. Consider the following example.

Let $A = \mathrm{diag}(1, \epsilon)$, $0 < \epsilon \leq 1$. Then $R = \mathrm{diag}(1, \epsilon)$, $W_R^{-1} Z_R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \epsilon & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Thus we obtain

$$\kappa_R(A) = \sqrt{1+\epsilon^2}, \quad \kappa_2(A) = 1/\epsilon,$$

and

$$\frac{\sqrt{2}\kappa_2(A)}{\kappa_R(A)} \sim \frac{\sqrt{2}}{\epsilon} \quad \text{as } \epsilon \to 0.$$

We see the first order perturbation bound (1.2) can severely overestimate the effect of a perturbation in A.

But it is possible that $\kappa_R(A)$ has the same order as $\kappa_2(A)$. If for example $A = \mathrm{diag}(\epsilon, 1)$ with $0 < \epsilon \leq 1$, then $R = \mathrm{diag}(\epsilon, 1)$, $W_R^{-1} Z_R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\epsilon & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Thus we get

$$\kappa_R(A) = \sqrt{1+1/\epsilon^2}, \quad \kappa_2(A) = 1/\epsilon,$$

and

$$\frac{\kappa_2(A)}{\kappa_R(A)} \to 1 \quad \text{as} \quad \epsilon \to 0.$$

Actually $\kappa_R(A)$ is the optimal measure of the sensitivity of the R factor of the QR factorization in the following sense when $m = n$. Suppose we are able to obtain a bound of the form

$$\frac{\|\dot{R}(0)\|_F}{\|R\|_2} \leq f(\|A\|_{2,F}, \|A^\dagger\|_{2,F}), \tag{2.12}$$

where f, a function of $\|A\|_2$ or $\|A\|_F$, $\|A^\dagger\|_2$ or $\|A^\dagger\|_F$, is some other measure of the sensitivity of the R factor of QR factorization than $\kappa_R(A)$. Solving (1.5) we have $\dot{R}(0) = \mathbf{T}_R^{-1}\hat{\mathbf{T}}_R Q_1^T G$. Noticing $\|Q_1^T G\|_F = \|G\|_F = \|A\|_2 = \|R\|_2$ when $m = n$, so that from (2.12) we get

$$\frac{\|\mathbf{T}_R^{-1}\hat{\mathbf{T}}_R Q_1^T G\|_F}{\|Q_1^T G\|_F} \leq f(\|A\|_{2,F}, \|A^\dagger\|_{2,F}).$$

As a consequence, we have for our operator norm

$$\|\mathbf{T}_R^{-1}\hat{\mathbf{T}}\|_F \leq f(\|A\|_{2,F}, \|A^\dagger\|_{2,F}),$$

so from (2.4),

$$\kappa_R(A) = \|W_R^{-1} Z_R\|_2 \leq f(\|A\|_{2,F}, \|A^\dagger\|_{2,F}).$$

By the above analysis, we can consider $\kappa_R(A)$ to be the normwise condition number for the R factor of the QR factorization.

Remark 2. $\kappa_R(A)$ can be estimated by the following power method.
Choose a unit vector v_0, and loop until convergence:
1. Solve $W_R u_k = Z_R v_{k-1}$ for u_k and $u_k := u_k/\|u_k\|_\infty$
2. Solve $W_R^T v_k = u_k$ for v_k and $v_k := Z_R v_k$
3. $\sigma_k = \|v_k\|_\infty$ and $v_k := v_k/\|v_k\|_\infty$

Upon convergence, σ_k is an estimate of $\|W_R^{-1} Z_R\|_2$. Note that the linear systems in innerstep 1 and 2 can be solved in $O(n^3)$ by using the special structure of W_R and Z_R.

Remark 3. Suppose the QR factorization of A is approached by using the standard column pivoting strategy: $AP = Q_1 R$, where P is an $n \times n$ permutation matrix designed so that the columns of A are interchanged, during the computation of the reduction, to make the leading diagonal elements of R as large as possible. Let the QR factorization of $(A + \Delta A)P$ be $(A + \Delta A)P = (Q_1 + \Delta Q_1)(R + \Delta R)$. Then by Theorem 2.1 we have

$$\frac{\|\Delta R\|_F}{\|R\|_2} \leq \kappa_R(AP)\epsilon + O(\epsilon^2),$$

and

$$1 \leq \kappa_R(AP) \leq \sqrt{2}\kappa_2(A).$$

Note that the first order bound (1.2) does not change when the QR factorization of A is approached by using any pivoting strategy. Clearly the perturbation bound (2.9) more closely reflects the structure of the problem. From the two examples in Remark 1, we can find that if we use the standard column pivoting for the second example, $\kappa_R(AP)$ will become the same as that for the first one, i.e. $\sqrt{1+\epsilon^2}$, showing how pivoting can improve the condition of the problem. In fact we have the following conclusion.

THEOREM 2.2. *Let $A \in \mathcal{R}^{m\times n}$ be of full column rank, with the QR factorization $AP = Q_1 R$ when the standard column pivoting is used. Then*

$$1 \leq \kappa_R(AP) = \|W_R^{-1} Z_R\|_F \leq \sqrt{\frac{1}{27}4^{n+1} + \frac{1}{3}n^2 + \frac{2}{9}n - \frac{4}{27}}. \tag{2.13}$$

There is a parametrized family of matrices $A(\theta)$, $\theta \in (0, \pi/2]$, for which

$$\|W_R^{-1} Z_R\|_F \to \sqrt{\frac{1}{27}4^{n+1} + \frac{1}{3}n^2 + \frac{2}{9}n - \frac{4}{27}} \quad as \quad \theta \to 0.$$

Proof. See Appendix. $\square$

Theorem 2.2 shows that when the standard column pivoting strategy is used $\kappa_R(AP)$ is bounded for fixed n no matter how large $\kappa_2(A)$ is. Many numerical experiments with the standard column pivoting strategy suggest that usually $\kappa_R(AP)$ is close to its lower bound of one.

Similar ideas apply to bounding the individual components of ΔR, as we now show.

THEOREM 2.3. *Let $A \in \mathcal{R}^{m\times n}$ be of full column rank, with the QR factorization $A = Q_1 R$, and let $|\Delta A| \le \epsilon C|A|$ for some nonnegative C with $c_{ij} \le 1$. If*

$$\epsilon \|C\|_2 \|A^\dagger\|_2 \|A\|_F < 1, \tag{2.14}$$

then $A + \Delta A$ has the QR factorization

$$A + \Delta A = (Q_1 + \Delta Q_1)(R + \Delta R),$$

where, using $\otimes$ to denote the Kronecker product,

$$\text{uvec}(|\Delta R|) \le \epsilon |W_R^{-1} Z_R||R^T \otimes I|\text{vec}(|Q_1^T|C|Q_1|) + O(\epsilon^2) \tag{2.15}$$

and

$$\frac{\|\Delta R\|_F}{\|R\|_F} \le \kappa_{R_c}(A)\eta\epsilon + O(\epsilon^2), \tag{2.16}$$

with $\kappa_{R_c}(A) \equiv \frac{\||W_R^{-1}Z_R||R^T\otimes I|\|_2}{\|R\|_F}$ and $\eta = \||Q_1^T|C|Q_1|\|_F$.

Proof. Let $G \equiv \Delta A/\epsilon$ (if $\epsilon = 0$ the theorem is trivial). If (2.14) holds, it is easy to show $A + tG$ has full column rank for all $t \in [0, \epsilon]$. Thus $A + \Delta A$ has a unique QR factorization, and as in the proof for Theorem 2.1, we have

$$\text{uvec}(\dot{R}(0)) = W_R^{-1} Z_R \text{vec}(Q_1^T \Delta A/\epsilon).$$

Thus

$$\begin{aligned}\text{uvec}(|\dot{R}(0)|) &\le |W_R^{-1} Z_R|\text{vec}(|Q_1^T \Delta A/\epsilon|) \\ &\le |W_R^{-1} Z_R|\text{vec}(|Q_1^T|C|Q_1||R|) \\ &\le |W_R^{-1} Z_R||R^T \otimes I|\text{vec}(|Q_1^T|C|Q_1|),\end{aligned}$$

so taking the 2-norm gives

$$\|\dot{R}(0)\|_F \le \||W_R^{-1} Z_R||R^T \otimes I|\|_2 \eta.$$

Thus (2.15) and (2.16) follow immediately from the Taylor series (1.10) for $R(t)$. $\square$

Remark 4. In Section 1, we mentioned that the condition number $\kappa_S(A)$ in [13] is independent of the column scaling of A or R. Now let us see if $\kappa_{R_c}(A)$ is also insensitive to such column scaling. For any nonsingular diagonal $\hat{D} = \operatorname{diag}(\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_n)$, let the QR factorization of $A\hat{D}^{-1}$ be $A\hat{D}^{-1} = Q_1 R\hat{D}^{-1}$ and let $\hat{R} = R\hat{D}^{-1}$. Substituting $R = \hat{R}\hat{D}$ into equation (2.1) gives

$$\hat{D}\hat{R}^T X + X^T \hat{R}\hat{D} = \hat{D}\hat{R}^T F + F^T \hat{R}\hat{D}.$$

Thus

$$\hat{R}^T X\hat{D}^{-1} + \hat{D}^{-1} X^T \hat{R} = \hat{R}^T F\hat{D}^{-1} + \hat{D}^{-1} F^T \hat{R},$$

which can be rewritten in the following matrix-vector form

$$W_{\hat{R}} \operatorname{uvec}(X\hat{D}^{-1}) = Z_{\hat{R}} \operatorname{vec}(F\hat{D}^{-1}).$$

Thus we have

$$W_{\hat{R}} \hat{D}_1^{-1} \operatorname{uvec}(X) = Z_{\hat{R}} \hat{D}_2^{-1} \operatorname{vec}(F),$$

where

$$\hat{D}_1 = \operatorname{diag}(\hat{d}_1, \underbrace{\hat{d}_2, \hat{d}_2}_{2}, \cdots, \underbrace{\hat{d}_n, \hat{d}_n, \cdots, \hat{d}_n}_{n}) \in \mathcal{R}^{\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}}, \tag{2.17}$$

$$\hat{D}_2 = \operatorname{diag}(\underbrace{\hat{d}_1, \hat{d}_1, \cdots, \hat{d}_1}_{n}, \underbrace{\hat{d}_2, \hat{d}_2, \cdots, \hat{d}_2}_{n}, \cdots, \underbrace{\hat{d}_n, \hat{d}_n, \cdots, \hat{d}_n}_{n}) \in \mathcal{R}^{n^2 \times n^2}. \tag{2.18}$$

Then it follows that

$$\operatorname{uvec}(X) = \hat{D}_1 W_{\hat{R}}^{-1} Z_{\hat{R}} \hat{D}_2^{-1} \operatorname{vec}(F).$$

Comparing this with (2.3) and noticing F is arbitrary, we must have

$$W_R^{-1} Z_R = \hat{D}_1 W_{\hat{R}}^{-1} Z_{\hat{R}} \hat{D}_2^{-1}.$$

Then

$$\begin{aligned} \kappa_{R_c}(A) &= \frac{\| |\hat{D}_1 W_{\hat{R}}^{-1} Z_{\hat{R}} \hat{D}_2^{-1}| |\hat{D}_2 (\hat{R}^T \otimes I)| \|_2}{\|\hat{R}\hat{D}\|_F} \\ &= \frac{\| |\hat{D}_1| |W_{\hat{R}}^{-1} Z_{\hat{R}}| |\hat{R}^T \otimes I| \|_2}{\|\hat{R}\hat{D}\|_F}. \end{aligned} \tag{2.19}$$

Usually

$$\| |\hat{D}_1| |W_{\hat{R}}^{-1} Z_{\hat{R}}| |\hat{R}^T \otimes I| \|_2 \approx \|\hat{D}_1\|_2 \| |W_{\hat{R}}^{-1} Z_{\hat{R}}| |\hat{R}^T \otimes I| \|_2, \quad \|\hat{R}\hat{D}\|_F \approx \|\hat{D}\|_2 \|\hat{R}\|_F.$$

Note that $\|\hat{D}_1\|_2 = \|\hat{D}\|_2$, so

$$\kappa_{R_c}(A) \approx \| |W_{\hat{R}}^{-1} Z_{\hat{R}}| |\hat{R}^T \otimes I| \|_2 / \|\hat{R}\|_F = \kappa_{R_c}(A\hat{D}^{-1}).$$

Therefore we can conclude that $\kappa_{R_c}(A)$ is insensitive to the column scaling of A. Now we use the example given by Zha [13] to illustrate this.

$$A = \begin{pmatrix} 1 & 10^{10} \\ 0 & 10^{10} \\ 1 & 10^{10} \end{pmatrix},$$

$$\kappa_2(A) = 2.1213e+10, \quad \kappa_S(A) = 3.1623e+00, \quad \kappa_{R_c}(A) = 1.8507e+00,$$

where $\kappa_S(A)$ takes the Frobenius norm in (1.11).

Remark 5. If we use the standard column pivoting in computing the QR factorization of A, $\kappa_{R_c}(AP)$ can be bounded in terms of n. In fact by Theorem 2.2

$$\kappa_{R_c}(AP) = \frac{\| |W_R^{-1} Z_R| |R^T \otimes I| \|_2}{\|R\|_F} \le \frac{\|W_R^{-1} Z_R\|_F \| |R^T| \|_2}{\|R\|_F}$$
$$\le \sqrt{\frac{1}{27}4^{n+1} + \frac{1}{3}n^2 + \frac{2}{9}n - \frac{4}{27}}.$$

However $\kappa_S(A)$ can be arbitrarily large, as shown by the following example

$$A = \begin{pmatrix} 2 & 1 \\ 0 & \epsilon \end{pmatrix} \quad \text{with very small } \epsilon,$$

$$\kappa_{R_c}(AP) \le 2, \quad \kappa_S(AP) = O(\frac{1}{\epsilon}) \quad \text{with the Frobenius norm.}$$

By the above analysis, we can consider $\kappa_{R_c}(A)$ to be the componentwise condition number for the R factor of the QR factorization.

3. Practical normwise bound and condition estimator. In Section 2 we derived new normwise and componentwise perturbation bounds for the R factor of the QR factorization and saw $\kappa_R(A)$ and $\kappa_{R_c}(A)$ are the normwise condition number and componentwise condition number for R, respectively. The drawback with those bounds are that $\kappa_R(A)$ and $\kappa_{R_c}(A)$ are both difficult to understand and (so far) to compute or estimate, except when we use pivoting, in which case $\kappa_R(AP)$ usually approaches its lower bound 1, also $\kappa_{R_c}(AP)$ approaches 1. For $\kappa_R(A)$ fortunately we can obtain an expression which, though it is large than $\kappa_R(A)$, is a good estimate of it, and also gives considerable insight into what is going on, and leads to an efficient and practical condition estimator even when we do not use pivoting.

With any nonsingular diagonal $D = \mathrm{diag}(d_1, d_2, \cdots, d_n)$ we can take $R = D\bar{R}$. It is easy to observe that

$$W_R^{-1} Z_R = D_1^{-1} W_{\bar{R}}^{-1} Z_{\bar{R}} D_2, \tag{3.1}$$

where

$$D_1 = \mathrm{diag}(d_1, \underbrace{d_1, d_2}_{2}, \cdots, \underbrace{d_1, d_2, \cdots, d_n}_{n}) \in \mathcal{R}^{\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}},$$

$$D_2 = \mathrm{diag}(\underbrace{d_1, d_2, \cdots, d_n}_{n}, \underbrace{d_1, d_2, \cdots, d_n}_{n}, \cdots, \underbrace{d_1, d_2, \cdots, d_n}_{n}) \in \mathcal{R}^{n^2 \times n^2}.$$

Let $A \circ B$ represent the Hadamard product of $A = (a_{ij}) \in \mathcal{R}^{m \times n}$ and $B = (b_{ij}) \in \mathcal{R}^{m \times n}$, that is $A \circ B = (a_{ij} b_{ij}) \in \mathcal{R}^{m \times n}$. Obviously we can write

$$D_1^{-1} W_{\bar{R}}^{-1} Z_{\bar{R}} D_2 \equiv W_{\bar{R}}^{-1} Z_{\bar{R}} \circ T, \tag{3.2}$$

where $T \in \mathcal{R}^{\frac{n(n+1)}{2} \times n^2}$ is

$$
\left[\begin{array}{ccccc|ccccc|c|ccccc}
1 & & & & & & & & & & & & & & & \\
\hline
 & \frac{d_2}{d_1} & & & & 1 & & & & & & & & & & \\
 & 1 & & & & & 1 & & & & & & & & & \\
\hline
 & \frac{d_2}{d_1} & \frac{d_3}{d_1} & & & & & & & & & & & & & \\
 & 1 & \frac{d_3}{d_2} & & & & & \frac{d_3}{d_2} & & & & & & & & \\
 & & 1 & & & & & 1 & & & & & & & & \\
\hline
 & \cdot & \cdot & \cdot & & & & \cdot & \cdot & \cdot & \cdot & & & & & \\
\hline
 & \frac{d_2}{d_1} & \frac{d_3}{d_1} & \cdot & \frac{d_n}{d_1} & & & & & & & 1 & & & & \\
 & 1 & \frac{d_3}{d_2} & \cdot & \frac{d_n}{d_2} & & & \frac{d_3}{d_2} & \cdot & \frac{d_n}{d_2} & & & 1 & & & \\
 & & \cdot & \cdot & \cdot & & & \cdot & \cdot & \cdot & & & & \cdot & & \\
 & & & 1 & \frac{d_n}{d_{n-1}} & & & & 1 & \frac{d_n}{d_{n-1}} & & & & & 1 & \\
 & & & & 1 & & & & & 1 & & & & & & 1
\end{array}\right],
$$

which has the same zero/nonzero structure as $W_{\bar{R}}^{-1} Z_{\bar{R}}$ and $W_R^{-1} Z_R$ (see Appendix A for the zero/nonzero structure of $W_R^{-1} Z_R$).

In order to estimate $\|W_{\bar{R}}^{-1} Z_{\bar{R}} \circ T\|_2$, we use the following lemma (see Theorem 5.5.3 in [4]).

Lemma 3.1. *Given any $A, B \in \mathcal{R}^{m\times n}$ we have*

$$\|A \circ B\|_2 \leq \min\{r(B), c(B)\} \|A\|_2,$$

where $r(B) \equiv \max_i (\sum_{j=1}^n |b_{ij}|^2)^{\frac{1}{2}}$ and $c(B) \equiv \max_j (\sum_{i=1}^m |b_{ij}|^2)^{\frac{1}{2}}$.

By the above lemma we have

$$\|W_{\bar{R}}^{-1} Z_{\bar{R}} \circ T\|_2 \leq \min\{r(T), c(T)\} \|W_{\bar{R}}^{-1} Z_{\bar{R}}\|_2, \tag{3.3}$$

where

$$r(T) = \max\{\sqrt{n}, \sqrt{j(\textstyle\sum_{i=j+1}^n (\frac{d_i}{d_j})^2 + 1)}, \quad j = 1, 2, \cdots, n-1\}, \tag{3.4}$$

and

$$c(T) = \max\{1, \sqrt{(n-j)(\textstyle\sum_{i=1}^j (\frac{d_{j+1}}{d_i})^2 + 1)}, \quad j = 1, 2, \cdots, n-1\}. \tag{3.5}$$

Combining (3.1), (3.2) and (3.3) with (2.6), we get

$$\kappa_R(A) = \|W_R^{-1} Z_R\|_2 \leq \sqrt{2} \min\{r(T), c(T)\} \kappa_2(\bar{R}) \equiv \tilde{\kappa}_R(A, D), \quad \text{say}, \tag{3.6}$$

and $\tilde{\kappa}_R(A, D)$ is a potential estimator for $\kappa_R(A)$. Since D is an arbitrary nonsingular diagonal matrix, it may be chosen to minimize $\tilde{\kappa}_R(A, D)$,

$$\tilde{\kappa}_R(A) \equiv \min_{\text{nonsingular diagonal } D} \tilde{\kappa}_R(A, D), \tag{3.7}$$

with the minimum being no greater than $\frac{n+1}{\sqrt{2}} \kappa_2(A)$ by taking $D = I$.

This approach provides another explanation as to why the standard column pivoting of A is so successful, making $\kappa_R(AP)$ approach its lower bound in nearly all cases. If A is ill-conditioned (so there is a large distance between the lower and upper bounds on $\kappa_R(A)$) and the QR factorization is computed with the standard pivoting, the ill-conditioning of A will usually reveal itself in the diagonal elements of R. Stewart [7] has shown that such upper triangular matrices are artificially ill-conditioned in the sense that they can be made well-conditioned by scaling the rows via D. Observing (3.4) and (3.5), we can expect $r(T)$ and $c(T)$ to be relatively small since pivoting is used, e.g., taking $d_i = r_{ii}$. This implies that $\tilde{\kappa}_R(AP, D)$, and therefore $\kappa_R(AP)$, will approach its lower bound. We can support this mathematically. In fact, if we take $D = \mathrm{diag}(r_{ii})$, then $d_1 \geq d_2 \geq \ldots \geq d_n$ when the standard column pivoting is used, so that $r(T), c(T) \leq \frac{n+1}{2}$. It is known that (see e.g. [1]) $\kappa_2(\bar{R}) \leq \sqrt{n(n+1)(4^n+6n-1)/18}$. Thus we have the following bound

$$\tilde{\kappa}_R(A, D) \leq (n+1)\sqrt{n(n+1)(4^n+6n-1)}/6.$$

The practical outcome of this analysis is that we now have an $O(n^2)$ condition estimator for the R factor of the QR factorization. By a well known result of van der Sluis [12], $\kappa_2(\bar{R})$ will be nearly optimal when the rows of $\bar{R}$ are equilibrated. Thus the procedure is to choose D in $R = D\bar{R}$ so that the rows of $\bar{R}$ are equilibrated, and use a condition estimator, see for example [3], to estimate $\kappa_2(\bar{R})$ in

$$1 \leq \kappa_R(A) \leq \tilde{\kappa}_R(A) \leq \tilde{\kappa}_R(A, D) = \sqrt{2}\min\{r(T), c(T)\}\kappa_2(\bar{R}). \tag{3.8}$$

We will use another approach to estimate $\kappa_{R_c}(A)$ in a coming paper.

4. Numerical experiments. In Section 2 we presented new first order normwise and componentwise perturbation bounds for the R factor of the QR factorization, defined $\kappa_R(A) \equiv \|W_R^{-1}Z_R\|_2$ and $\kappa_{R_c}(A) \equiv \||W_R^{-1}Z_R||R^T \otimes I|\|_2/\|R\|_F$ as the normwise condition number and componentwise condition number for the R factor, respectively. Our new first order results generally are sharper than the existing results. In Section 3, a practical normwise condition estimator $\tilde{\kappa}_R(A, D) = \sqrt{2}\min\{r(T), c(T)\}\kappa_2(\bar{R})$ was given. Now we use some numerical tests to illustrate our results and analyses.

We give one set of $n \times n$ Pascal matrices ($n = 1, 2, \ldots, 15$) as test examples. The normwise results are shown in Table 1. In this table, $R = D\bar{R}$, where $D = \mathrm{diag}(d_{ii})$ with $d_{ii} = \sqrt{\sum_{j=i}^n r_{ij}^2}$. The componentwise results are shown in Table 2.

Note in Table 1 how $\sqrt{2}\kappa_2(A)$ can be far worse than the optimal condition number $\kappa_R(A)$. Pivoting is seen to give a significant improvement on $\kappa_R(A)$, bringing $\kappa_R(AP)$ very close to its lower bound 1. Also we observe that $\tilde{\kappa}_R(A, D)$ is a very good estimate for $\kappa_R(A)$, whether we use pivoting or not. The results shown in Table 1 suggest the ratio $\kappa_R(A, D)/\kappa_R(A)$ may be bounded by a reasonable function of n.

The results in Table 2 show $\kappa_S(A)$ in [13] can be significantly larger than our $\kappa_{R_c}(A)$. Again pivoting gives further improvement on $\kappa_{R_c}(A)$. Even though $\kappa_S(A)$ does change when pivoting is used, the change is not significant. We also observe that our new componentwise condition number is better (smaller) than the corresponding new normwise condition number. It should be, since the componentwise perturbation in Theorem 1.2 has special structure, while no structure is imposed on the normwise perturbation.

Table 1

n	$\kappa_R(A)$	$\tilde{\kappa}_R(A, D)$	$\kappa_R(AP)$	$\tilde{\kappa}_R(AP, D)$	$\sqrt{2}\kappa_2(A)$
1	1.0000e+00	1.4142e+00	1.0000e+00	1.4142e+00	1.4142e+00
2	1.8708e+00	4.8471e+00	1.1832e+00	2.5342e+00	9.6932e+00
3	4.6332e+00	2.8659e+01	1.2892e+00	4.3251e+00	8.7658e+01
4	1.4359e+01	1.3865e+02	1.6901e+00	8.1713e+00	9.7855e+02
5	4.9822e+01	6.0882e+02	1.8129e+00	1.1443e+01	1.2046e+04
6	1.8021e+02	2.6285e+03	2.2296e+00	1.4413e+01	1.5668e+05
7	6.6572e+02	1.1215e+04	2.0543e+00	1.6725e+01	2.1120e+06
8	2.4916e+03	4.7441e+04	2.6264e+00	2.3226e+01	2.9197e+07
9	9.4091e+03	1.9936e+05	3.4892e+00	3.2789e+01	4.1123e+08
10	3.5768e+04	8.3338e+05	3.3983e+00	3.8082e+01	5.8763e+09
11	1.3668e+05	3.4692e+06	3.3616e+00	3.9552e+01	8.4943e+10
12	5.2442e+05	1.4391e+07	3.3468e+00	4.2555e+01	1.2394e+12
13	2.0190e+06	5.9523e+07	3.3417e+00	5.1420e+01	1.8226e+13
14	7.7958e+06	2.4558e+08	3.6106e+00	5.4381e+01	2.6979e+14
15	3.0179e+07	1.0111e+09	3.3430e+00	5.1780e+01	4.0149e+15

Table 2

n	$\kappa_{R_c}(A)$	$\kappa_S(A)$	$\kappa_{R_c}(AP)$	$\kappa_S(AP)$
1	1.0000e+00	1.0000e+00	1.0000e+00	1.0000e+00
2	1.6555e+00	7.0000e+00	1.0649e+00	7.0000e+00
3	2.3867e+00	5.4536e+01	1.0445e+00	5.1900e+01
4	3.4644e+00	4.8324e+02	1.0371e+00	4.4853e+02
5	5.2564e+00	4.6197e+03	1.0241e+00	4.3410e+03
6	9.1407e+00	4.6194e+04	1.0190e+00	4.2220e+04
7	1.6100e+01	4.7551e+05	1.0144e+00	4.2823e+05
8	2.8450e+01	5.0655e+06	1.0132e+00	4.4331e+06
9	5.2951e+01	5.4834e+07	1.5649e+00	4.7457e+07
10	1.0047e+02	5.9765e+08	1.4769e+00	5.0896e+08
11	1.8884e+02	6.5520e+09	1.4277e+00	5.4532e+09
12	3.5867e+02	7.2190e+10	1.3981e+00	5.8679e+10
13	6.9278e+02	7.9883e+11	1.3784e+00	6.4313e+11
14	1.3326e+03	8.8732e+12	1.3645e+00	7.0468e+12
15	2.5614e+03	9.8858e+13	1.3542e+00	7.7607e+13

5. Summary and conclusions. We have presented new normwise and componentwise perturbation analyses for the R factor of the QR factorization.

Although the Stewart and Sun first order normwise perturbation bound (1.2) is relevant in the sense that some problems do attain close to the indicated condition, we have shown that it gives a large over-bound for most problems. The more refined bound (2.9) is usually significantly stronger, never weaker, and the resulting condition number $\kappa_R(A)$ more accurately reflects the true sensitivity of the problem. Further, the sizes of our condition numbers depend on any column pivoting used, and can be bounded in terms of n when the standard column pivoting is use, no matter how large $\kappa_2(A)$ is. Numerical results suggest that the standard column pivoting strategy leads to a near optimally conditioned factorization for R with a given A in $AP = Q_1R$.

Because of the difficulty in computing $\kappa_R(A)$, there was need for, and fortunately we have been able to give a good estimate $\tilde{\kappa}_R(A, D)$. Although the new estimate is somewhat weaker, it can be estimated by the usual condition estimators (see, for example, [3]) in $O(n^2)$.

Although the first order componentwise perturbation bound (1.12) given by Zha [13] is relevant in the sense that it is invariant under column scaling, it also gives a large over-bound for most problems. Pivoting cannot change it much. Numerical tests and analyses suggest the new bound (2.15) usually is stronger. The resulting componentwise condition number $\kappa_{R_c}(A)$ is strongly effected by pivoting, and the size is bounded for fix n when the standard column pivoting is used, while $\kappa_S(A)$ can be arbitrarily large.

All these results are being rewritten as a paper using the "up"-notation of Stewart [8], see also Chang, Paige and Stewart [1].

Appendix:

A. A bound for $\|W_R^{-1} Z_R\|_F$ when the standard column pivoting is used. If we use the standard pivoting strategy in computing the QR factorization of matrix $A \in \mathcal{R}^{m \times n}$ with rank n, then the matrix R satisfies

$$r_{kk}^2 \geq \sum_{i=k}^{j} r_{ij}^2, \quad j = k+1, \ldots, n, \quad k = 1, \ldots, n. \tag{A.1}$$

We first introduce some notation. For any vector c_j, denote by $c_j^{(i)}$ the vector of the leading i elements of c_j. For any matrix $C \in \mathcal{R}^{m \times n}$, denote the submatrix of C that lies successively in row i till row i' and column j till column j' as $C_{([i,i'],[j,j'])}$, or $C_{[i,i']}$ if $i = j$ and $i' = j'$. Denote by $e_j^{[i]}$ the jth column of an $i \times i$ unit matrix. Denote the $n \times n$ identity matrix by I_n.

Write $H = W_R^{-1} Z_R$, and partition H conformably with Z_R (each block H_{ij} is $i \times n$). Let $y \in \mathcal{R}^{\frac{n(n+1)}{2}}$ be the k-th column of the j-th block column of H. Write $y = (y_1, y_2, \ldots, y_i, \ldots, y_n)^T$, where $y_i = (y_{i1}, y_{i2}, \ldots, y_{ii})^T \in \mathcal{R}^i$, according to the partition of H.

It is easy to observe that the first j columns of the j-th block column of Z_R are just the j-th block column of W_R. Thus

$$H_{ii[1,i]} = I_i, \quad H_{ij([1,i],[1,j])} = 0, \quad i \neq j, \quad i, j = 1, \ldots, n. \tag{A.2}$$

We assume $k > j$ from now on. The k-th column of the j-th block column of Z_R

$$z = (\underbrace{0, 0, \ldots, 0}_{\frac{(k-1)k}{2}}, r_{kk} e_j^{[k]T}, r_{k,k+1} e_j^{[k+1]T}, \ldots, r_{ki} e_j^{[i]T}, \ldots, r_{kn} e_j^{[n]T})^T,$$

satisfies

$$W_R y = z. \tag{A.3}$$

It follows from (A.3) that

$$y_i = 0, \quad i = 1, \ldots, k-1, \tag{A.4}$$

and for $k+1 \le i \le n$,

$$\text{(A.5)}\quad \begin{bmatrix} R_{[1,k]}^T & & & & \\ e_k^{[k+1]} r_{k+1}^{(k)\,T} & R_{[1,k+1]}^T & & & \\ \cdot & \cdot & \cdot & & \\ e_k^{[l]} r_l^{(k)T} & e_{k+1}^{[l]} r_l^{(k+1)T} & \cdot & R_{[1,l]}^T & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ e_k^{[i]} r_i^{(k)T} & e_{k+1}^{[i]} r_i^{(k+1)T} & \cdot & e_l^{[i]} r_i^{(l)T} & \cdot \quad R_{[1,i]}^T \end{bmatrix} \begin{bmatrix} y_k \\ y_{k+1} \\ \cdot \\ y_l \\ \cdot \\ y_i \end{bmatrix} = \begin{bmatrix} r_{kk} e_j^{[k]} \\ r_{k,k+1} e_j^{[k+1]} \\ \cdot \\ r_{kl} e_j^{[l]} \\ \cdot \\ r_{ki} e_j^{[i]} \end{bmatrix}.$$

The first block row of equations (A.5) shows

$$\text{(A.6)}\qquad y_{k1} = y_{k2} = \ldots = y_{k,j-1} = 0,$$

$$\text{(A.7)}\qquad R_{[j,k-1]}^T \begin{bmatrix} y_{kj} \\ y_{k,j+1} \\ \cdot \\ y_{k,k-1} \end{bmatrix} = r_{kk} e_1^{[k-j]}$$

and (using (A.6) as well)

$$\text{(A.8)}\qquad r_{jk} y_{kj} + r_{j+1,k} y_{k,j+1} + \cdots + r_{k-1,k} y_{k,k-1} + r_{kk} y_{kk} = 0.$$

Now from (A.7) we have

$$\text{(A.9)}\qquad (y_{kj}, y_{k,j+1}, \ldots y_{k,k-1})^T = r_{kk} R_{[j,k-1]}^{-T} e_1^{[k-j]},$$

so from (A.8), we get

$$\text{(A.10)}\qquad y_{kk} = -(r_{jk}, r_{j+1,k}, \ldots, r_{k-1,k}) R_{[j,k-1]}^{-T} e_1^{[k-j]}.$$

Now we obtain y_i, $i = k+1, \ldots, n$. From the first k equations of the last block row of (A.5), it follows

$$\text{(A.11)}\qquad y_{i1} = y_{i2} = \ldots = y_{i,j-1} = 0,$$

$$\text{(A.12)}\qquad R_{[j,k-1]}^T \begin{bmatrix} y_{ij} \\ y_{i,j+1} \\ \cdot \\ y_{i,k-1} \end{bmatrix} = r_{ki} e_1^{[k-j]},$$

and (using (A.6) and (A.11) as well)

$$\text{(A.13)}\qquad (r_{ji}, \ldots, r_{k-1,i}, r_{ki}) \begin{bmatrix} y_{kj} \\ \vdots \\ y_{k,k-1} \\ y_{kk} \end{bmatrix} + (r_{jk}, \ldots, r_{k-1,k}, r_{kk}) \begin{bmatrix} y_{ij} \\ \vdots \\ y_{i,k-1} \\ y_{ik} \end{bmatrix} = 0.$$

Next from (A.12), we get

$$\text{(A.14)}\qquad (y_{ij}, y_{i,j+1}, \ldots, y_{i,k-1})^T = r_{ki} R_{[j,k-1]}^{-T} e_1^{[k-j]}.$$

Using (A.9), (A.10) and (A.14), we get, from (A.13)

$$(r_{ji}, r_{j+1,i}, \ldots, r_{k-1,i}) r_{kk} R_{[j,k-1]}^{-T} e_1^{[k-j]} - r_{ki}(r_{jk}, r_{j+1,k}, \ldots, r_{k-1,k}) R_{[j,k-1]}^{-T} e_1^{[k-j]}$$
$$+ (r_{jk}, r_{j+1,k}, \ldots, r_{k-1,k}) r_{ki} R_{[j,k-1]}^{-T} e_1^{[k-j]} + r_{kk} y_{ik} = 0.$$

Thus,

$$y_{ik} = -(r_{ji}, r_{j+1,i}, \ldots, r_{k-1,i}) R_{[j,k-1]}^{-T} e_1^{[k-j]}. \tag{A.15}$$

Notice if we take $i = k$ in (A.11), (A.14) and (A.15), then the corresponding equalities (A.6), (A.9) and (A.10) are obtained, respectively. Thus (A.11), (A.14) and (A.15) actually holds for all $i \geq k$.

Write $D = \mathrm{diag}(r_{jj}, r_{j+1,j+1}, \cdots, r_{k-1,k-1})$, and define

$$G = R_{[j,k-1]}^{T} D^{-1}.$$

In view of the inequalities (A.1), the elements of $G = (g_{pq})$ satisfy

$$g_{pp} = 1, \quad g_{pq} \leq 1 \quad \text{for } p > q, \quad p = 1, \ldots, k-j.$$

It is easy to show

$$|G^{-1} e_1^{[k-j]}| \leq (1, 1, 2, \ldots, 2^{k-j-2})^T. \tag{A.16}$$

Thus,

$$|R_{[j,k-1]}^{-T} e_1^{[k-j]}| = |D^{-1} G^{-1} e_1^{[k-j]}| \leq (\frac{1}{r_{jj}}, \frac{1}{r_{j+1,j+1}}, \frac{2}{r_{j+2,j+2}}, \ldots, \frac{2^{k-j-2}}{r_{k-1,k-1}})^T, \tag{A.17}$$

so that from (A.14) and (A.15) it follows

$$|y_{ij}| \leq \frac{|r_{ki}|}{r_{jj}}, \quad |y_{it}| \leq 2^{t-j-1} \frac{|r_{ki}|}{r_{tt}}, \quad t = j+1, j+2, \ldots, k-1, \tag{A.18}$$

and

$$\begin{aligned} |y_{ik}| &= |-(r_{ji}, r_{j+1,i}, \ldots, r_{k-1,i}) R_{[j,k-1]}^{-T} e_1^{[k-j]}| \\ &\leq \frac{|r_{ji}|}{r_{jj}} + \sum_{t=j+1}^{k-2} 2^{t-j-1} \frac{|r_{ti}|}{r_{tt}} + 2^{k-j-2} \frac{|r_{k-1,i}|}{r_{k-1,k-1}} \\ &\leq 1 + \sum_{t=j+1}^{k-2} 2^{t-j-1} + 2^{k-j-2} \frac{|r_{k-1,i}|}{r_{k-1,k-1}} \quad \text{(using (A.1))} \\ &\leq 2^{k-j-2}(1 + \frac{|r_{k-1,i}|}{r_{k-1,k-1}}). \end{aligned} \tag{A.19}$$

Now we would like to show

$$y_{i,k+1} = y_{i,k+1} = \ldots = y_{i,i} = 0 \tag{A.20}$$

by induction on i with $i \geq k+1$. By (A.14) and (A.15), for any i with $k \leq i \leq n$, we have

$$(\text{A.21}) \qquad (r_{ji}, \cdots, r_{k-1,i}, r_{ki}) \begin{bmatrix} y_{ij} \\ \vdots \\ y_{i,k-1} \\ y_{ik} \end{bmatrix}$$
$$= (r_{ji}, \cdots, r_{k-1,i}) R_{[j,k-1]}^{-T} r_{ki} e_1^{[k-j]} - r_{ki}(r_{ji}, \cdots, r_{k-1,i}) R_{[j,k-1]}^{-T} e_1^{[k-j]}$$
$$= 0.$$

Thus, when $i = k+1$,

$$(\text{A.22}) \qquad \sum_{t=j}^{k} r_{t,k+1} y_{k+1,t} = 0.$$

Notice the last equation of the second block row of (A.5) is

$$\sum_{t=1}^{j-1} r_{t,k+1} y_{k+1,t} + \sum_{t=j}^{k} r_{t,k+1} y_{k+1,t} + r_{k+1,k+1} y_{k+1,k+1} = 0,$$

which, together with (A.11) with $i = k+1$ and (A.22), gives

$$y_{k+1,k+1} = 0.$$

Suppose $y_{l,k+1} = y_{l,k+1} = \ldots = y_{l,l} = 0$ is true for all l with $k+1 \leq l \leq i-1$. Now we show (A.20) is also true. Using (A.14) and (A.15) with $i = l$ or directly, we have, for $l = k+1, \ldots, i-1$,

$$(\text{A.23}) \qquad (r_{ji}, \cdots, r_{k-1,i}, r_{ki}) \begin{bmatrix} y_{lj} \\ \vdots \\ y_{l,k-1} \\ y_{l,k} \end{bmatrix} + (r_{jl}, \cdots, r_{k-1,l}, r_{kl}) \begin{bmatrix} y_{ij} \\ \vdots \\ y_{i,k-1} \\ y_{i,k} \end{bmatrix}$$
$$= (r_{ji}, \cdots, r_{k-1,i}) r_{kl} R_{[j,k-1]}^{-T} e_1^{[k-j]} - r_{ki}(r_{jl}, \cdots, r_{k-1,l}) R_{[j,k-1]}^{-T} e_1^{[k-j]}$$
$$+ (r_{jl}, \cdots, r_{k-1,l}) r_{ki} R_{[j,k-1]}^{-T} e_1^{[k-j]} - r_{kl}(r_{ji}, \cdots, r_{k-1,i}) R_{[j,k-1]}^{-T} e_1^{[k-j]}$$
$$= 0.$$

Notice the last $i-k$ equations of the last block row of (A.5) are

$$\sum_{t=1}^{j-1} r_{ti} y_{lt} + \sum_{t=j}^{k} r_{ti} y_{lt} + \sum_{t=k+1}^{l} r_{ti} y_{lt} + \sum_{t=1}^{j-1} r_{tl} y_{it} + \sum_{t=j}^{k} r_{tl} y_{it} + \sum_{t=k+1}^{l} r_{tl} y_{it} = 0$$
$$l = k+1, \ldots, i-1,$$

$$\sum_{t=1}^{j-1} r_{ti} y_{it} + \sum_{t=j}^{k} r_{ti} y_{it} + \sum_{t=k+1}^{i} r_{ti} y_{it} = 0,$$

which, together with (A.11), (A.21), (A.23) and the induction hypothesis, gives

$$\sum_{t=k+1}^{l} r_{tl} y_{it} = 0, \quad l = k+1, \ldots, i.$$

Thus we deduce (A.20) from this triangular system.

So far we have obtained the value or the bound of each elements of H. We now give a bound on $\|H\|_F$. First we consider the (i,j) block H_{ij}, $i > j$. From (A.2), (A.4) and (A.11)), we know the first j columns, the last $n-i$ column and the first $j-1$ rows of H_{ij} are zero, respectively. So we just consider the remaining part $H_{ij_{([j,i],[j+1,i])}}$. Using (A.18) and (A.19), we have

$$
\begin{array}{l}
|H_{ij_{([j,i],[j+1,i])}}| \\
\le \begin{array}{c}
\begin{array}{ccccc} \quad j+1 \quad & \quad j+2 \quad & \quad & \qquad i-1 \qquad & \qquad i \qquad \end{array} \\
\left[\begin{array}{ccccc}
\frac{|r_{j+1,i}|}{r_{jj}} & \frac{|r_{j+2,i}|}{r_{jj}} & \cdot & \frac{|r_{i-1,i}|}{r_{jj}} & \frac{r_{ii}}{r_{jj}} \\
\frac{|r_{ji}|}{r_{jj}} & \frac{|r_{j+2,i}|}{r_{j+1,j+1}} & \cdot & \frac{|r_{i-1,i}|}{r_{j+1,j+1}} & \frac{r_{ii}}{r_{j+1,j+1}} \\
 & 1 + \frac{|r_{j+1,i}|}{r_{j+1,j+1}} & \cdot & 2\frac{|r_{i-1,i}|}{r_{j+2,j+2}} & 2\frac{r_{i,i}}{r_{j+2,j+2}} \\
 & & \cdot & \cdot & \cdot \\
 & & & 2^{i-j-3}(1 + \frac{|r_{i-2,i}|}{r_{i-2,i-2}}) & 2^{i-j-2}\frac{r_{i,i}}{r_{i-1,i-1}} \\
 & & & & 2^{i-j-2}(1 + \frac{|r_{i-1,i}|}{r_{i-1,i-1}})
\end{array}\right]
\end{array}
\begin{array}{c} \\ j \\ j+1 \\ j+2 \\ \\ i-1 \\ i \end{array} .
\end{array}
$$

Using (A.1), it follows

$$
\begin{aligned}
\text{(A.24)} \qquad \|H_{ij}\|_F^2 &= \|H_{ij_{([j,i],[j+1,i])}}\|_F^2 \\
&= (\sum_{k=j+1}^{i} \frac{|r_{ki}|^2}{r_{jj}^2} + \frac{|r_{ji}|^2}{r_{jj}^2}) + (\sum_{k=j+2}^{i} \frac{|r_{ki}|^2}{r_{j+1,j+1}^2} + (1 + \frac{|r_{j+1,i}|}{r_{j+1,j+1}})^2) \\
&\quad + (\sum_{k=j+3}^{i} 2^2 \frac{|r_{ki}|^2}{r_{j+2,j+2}^2} + (2^2(1 + \frac{|r_{j+2,i}|}{r_{j+2,j+2}})^2) + \cdots \\
&\quad + 2^{2(i-j-2)} \frac{r_{i,i}^2}{r_{i-1,i-1}^2} + 2^{2(i-j-2)}(1 + \frac{|r_{i-1,i}|}{r_{i-1,i-1}})^2) \\
&\le 1 + (1+3) + (2^2 + 3 \cdot 2^2) + \cdots + ((2^{i-j-2})^2 + 3 \cdot (2^{i-j-2})^2) \\
&= (4^{i-j} - 1)/3.
\end{aligned}
$$

Thus

$$
\begin{aligned}
\|H\|_F^2 &= \sum_{i=1}^{n} \|H_{ii}\|_F^2 + \sum_{i=2}^{n} \sum_{j=1}^{i-1} \|H_{ij}\|_F^2 \\
&\le \sum_{i=1}^{n} i + \sum_{i=2}^{n} \sum_{j=1}^{i-1} (4^{i-j} - 1)/3 \\
&= \frac{1}{27} 4^{n+1} + \frac{1}{3} n^2 + \frac{2}{9} n - \frac{4}{27},
\end{aligned}
$$

so that we have

$$\|H\|_F \leq \sqrt{\frac{1}{27}4^{n+1} + \frac{1}{3}n^2 + \frac{2}{9}n - \frac{4}{27}}. \tag{A.25}$$

Now we would like to show the following $n \times n$ Kahan matrix

$$R(\theta) = \mathrm{diag}(1, s, \ldots, s^{n-1}) \begin{bmatrix} 1 & -c & -c & \cdot & -c \\ & 1 & -c & \cdot & -c \\ & & 1 & \cdot & -c \\ & & & \cdot & \cdot \\ & & & & 1 \end{bmatrix} \tag{A.26}$$

with $c = \cos(\theta)$ and $s = \sin(\theta)$ can make $||H||_F$ approximate the bound of (A.25) when θ tends to zero.

Clearly $R(\theta)$ satisfies (A.1). Through simple computations, from (A.14) and (A.15), we get the nonzero part of the k-th column of the j-th block column of $H(\theta) = W_{R(\theta)}^{-1} Z_{R(\theta)}$ with $k > j$ as follows.

$$y_{kj}(\theta) = -s^{k-j}, \quad y_{kt}(\theta) = -s^{k-t}c(1+c)^{t-j-1}, \quad t = j+1, j+2, \ldots, k-1,$$

$$y_{ij}(\theta) = -s^{k-j}c, \quad y_{it}(\theta) = -s^{k-t}c^2(1+c)^{t-j-1}, \quad t = j+1, \ldots, k-1, \quad i = k+1, \ldots, n,$$

$$y_{ik} = c(1+c)^{k-j-1}, \quad i = k, k+1, \ldots, n.$$

Thus

$$\lim_{\theta \to 0} \|H_{ij}(\theta)\|_F^2 = \lim_{\theta \to 0} \sum_{k=j+1}^{i} (c(1+c)^{k-j-1})^2 = (4^{i-j} - 1)/3.$$

That is to say $\|H_{ij}(\theta)\|^2$ will converge to the upper bound (A.24) when θ tends to zero. Therefore the conclusion follows.

Acknowledgement. We would like to thank Ji-guang Sun for his suggestions and encouragement, and for providing us with draft versions of his work.

REFERENCES

[1] X. Chang, C.C. Paige and G.W. Stewart, *New perturbation analyses for the Cholesky factorization.* IMA J. Numer. Anal., to appear.

[2] G.H. Golub and C.F. Van Loan, *Matrix Computations*, The Johns Hopkins University press, 2nd ed., Baltimore, MD, 1989.

[3] N. J. Higham, *A survey of condition number estimation for triangular matrices*, *SIAM Rev.* **29** (1987), 575–596

[4] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, New York, 1991.

[5] G.W. Stewart, *Perturbation bounds for the QR factorization of a matrix*, SIAM J. Numer. Anal., 14 (1977), 509–518.

[6] G.W. Stewart, *On the perturbation of LU, Cholesky, and QR factorizations*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 1141–1146.

[7] G.W. Stewart, *The triangular matrices of Gaussian elimination and related decompositions*, Technical Report, CS-TR-3533 UMIACS-TR-95-91, University of Maryland, Department of Computer Science, 1995.

[8] G.W. Stewart, *On the Perturbation of LU and Cholesky Factors*, Technical Report, CS-TR-3535 UMIACS-TR-95-93, University of Maryland, Department of Computer Science, 1995.

[9] G.W. Stewart and J.-G. Sun, *Matrix perturbation theory*, Academic Press, London, 1990.

[10] J.-G. Sun, *Perturbation bounds for the Cholesky and QR factorization*, BIT, 31 (1991), pp. 341–352.

[11] J.-G. Sun, *Componentwise perturbation bounds for some matrix decompositions*, BIT, 32 (1992), pp. 702–714.

[12] A. van der Sluis, *Condition numbers and equilibration of matrices*, Numerische Mathematik, 14 (1969) 14–23.

[13] H. Zha, *A componentwise perturbation analysis of the QR decomposition*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 1124-1131.