A note on the stability of the LU factorization of Hessenberg matrices

C. Brittin and M. I. Bueno

Abstract. In this paper we show that Doolittle’s method to compute the LU factorization of Hessenberg matrices is mixed forward-backward stable and therefore, componentwise forward stable. We also conjecture that this algorithm for computing the LU factorization of dense matrices is forward stable.

Keywords. LU factorization, Hessenberg matrices, mixed forward-backward stable algorithm, forward stability.

1. Introduction

Let \( H \) be any \( n \)-by-\( n \) matrix. This matrix is said to have an LU factorization if there exist a lower triangular matrix \( L \) and an upper triangular matrix \( U \) such that \( H = LU \). The LU factorization is one of the more important factorizations in Matrix Analysis and Numerical Analysis. It is well known \([4, 5]\) that the Doolittle’s method for computing the LU factorization of dense matrices is neither backward stable nor stable in the mixed forward-backward sense \([5]\). However, when special structured matrices are considered, more can be said about the stability of this factorization. For instance, the LU factorization of nonsingular, tridiagonal and diagonally dominant by rows or column matrices is backward stable \([5]\) and the LU factorization of tridiagonal matrices is stable in the mixed forward-backward sense \([1]\).

In this paper we consider Hessenberg matrices. A matrix \( H \) is said to be upper (lower) Hessenberg if \( H(i, j) = 0 \) for \( i - j > 1, \ (j - i > 1) \). In the sequel we will only consider upper Hessenberg matrices since all the results are also valid for lower Hessenberg matrices. In terms of stability, it is known \([5]\) that the growth factor for

This research has been supported by the Ministerio de Ciencia y Tecnología of Spain through grant BFM2003-06335-C03-02 and the Ministerio de Educación y Ciencia of Spain through the postdoctoral fellowship EX2004-0658.
the LU factorization with pivoting of $n$-by-$n$ Hessenberg matrices is bounded by $n$. This implies that the LU factorization with partial pivoting of Hessenberg matrices is backward stable. We have not found any formal results about the stability of the LU factorization without pivoting of Hessenberg matrices. In this paper we prove that the LU factorization of Hessenberg matrices is stable in the mixed forward-backward sense. As a consequence, we show that, for this special type of matrices, that factorization is also componentwise forward stable, i.e., the obtained forward errors are of similar magnitude to those produced by a backward stable algorithm. The stability of the LU factorization without pivoting is important in those problems in which the structure of the original matrix must be preserved (e.g. [2]). Notice that pivoting does not preserve the structure of matrix problems in general.

Based on numerical experiments, in this paper we also conjecture that the LU factorization of dense matrices is componentwise forward stable.

2. Stability of the LU factorization of Hessenberg matrices

Let $H$ be any $n$-by-$n$ upper Hessenberg matrix,

$$H = \begin{bmatrix}
h_{11} & h_{12} & h_{13} & \cdots & h_{1,n-1} & h_{1n} \\
h_{21} & h_{22} & h_{23} & \cdots & h_{2,n-1} & h_{2n} \\
0 & h_{32} & h_{33} & \cdots & h_{3,n-1} & h_{3n} \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & h_{n,n-1} & h_{nn}
\end{bmatrix}.$$ 

Let us assume that the first $n - 1$ leading principal minors of $H$ are nonzero. Then there exists an LU factorization of $H$. Let $H = LU$ be the unique LU factorization of $H$, where $L$ is a unit lower triangular matrix. Notice that $L$ is also a bidiagonal matrix. In the sequel, we use the following notation for the entries of the factors $L$ and $U$:

$$L = \begin{bmatrix}
1 & 0 & \cdots & 0 & 0 \\
l_1 & 1 & \cdots & 0 & 0 \\
l_2 & & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
l_{n-1} & & & \cdots & 1
\end{bmatrix}, \quad U = \begin{bmatrix}
u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\
u_{22} & u_{23} & \cdots & u_{2n} \\
u_{33} & \cdots & u_{3n} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
u_{nn} & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}.$$ 

The following MATLAB code computes the unique LU factorization of $H$.

**Algorithm 2.1.** Let $H$ be an $n$-by-$n$ upper Hessenberg matrix. This algorithm computes the unique LU factorization of $H$.

- $u_{ii} = h_{ii}$, for $i = 1, \ldots, n$ 
- for $i = 1 : n - 1$ 
- $l_i = h_{i+1,i} / u_{ii}$; 
- for $j = i + 1 : n$
The computational cost of Algorithm 2.1 is $n^2 + 2n - 1$ flops.

Next we present the backward error analysis of Algorithm 2.1. It has been done using the standard model of floating point arithmetic [5]:

$$fl(x \text{ op } y) = (x \text{ op } y)(1 + \delta) = x \text{ op } y(1 + \eta),$$

where $x$ and $y$ are floating point numbers, $\text{op} = +, -, \ast, /$, and $\epsilon$ is the unit roundoff of the machine. From now on, given a vector $v$ and a matrix $A$, $|v|$ and $|A|$ denote, respectively the vector and the matrix whose entries are the absolute values of the entries of $v$ and $A$.

We assume that the input parameters $h_{ij}$, with $i - j \leq 1$, are affected by small relative errors $(1 + \epsilon h_{ij})$, where $\max\{|\epsilon h_{ij}|\} \leq \epsilon$. These errors on the inputs may come from rounding errors committed in storing them in the computer.

**Theorem 2.2.** Let $H$ be an $n \times n$ upper Hessenberg matrix with an LU factorization.

Let $H = LU$ be the unique LU factorization of $H$. If $\hat{L}, \hat{U}$ are the factors obtained by applying Algorithm 2.1 to the matrix $\hat{H}$ with floating entries $\hat{h}_{ij}$ where

$$\hat{h}_{ij} = h_{ij}(1 + \epsilon h_{ij}), \quad i \leq j + 1, \quad j = 1, \ldots, n, \quad |\epsilon h_{ij}| \leq \epsilon$$

then, there exists an upper Hessenberg matrix $\Delta H$ such that

$$H + \Delta H = \hat{L}\hat{U}, \quad |\Delta H| \leq (2\epsilon + \epsilon^2)\left[|H| + |\hat{U}|\right].$$

**Proof.** For the computed quantities,

$$\hat{h}_i = \frac{h_{i+1,j}(1 + \epsilon h_{i+1,j})(1 + \eta_i)}{\hat{u}_{ii}}, \quad |\epsilon h_{i+1,j}|, |\eta_i| \leq \epsilon$$

and we get

$$|h_{i+1,i} - \hat{h}_i| \leq (2\epsilon + \epsilon^2)|h_{i+1,i}|.$$  

On the other hand, for $j = i + 1, \ldots, n$,

$$\hat{u}_{i+1,j}(1 + \beta_{i+1,j}) = h_{i+1,j}(1 + \epsilon h_{i+1,j}) - \frac{\hat{h}_i \hat{u}_{ij}}{1 + \lambda_{i+1,j}}, \quad |\beta_{i+1,j}|, |\epsilon h_{i+1,j}|, |\lambda_{i+1,j}| \leq \epsilon,$$

or equivalently,

$$\hat{u}_{i+1,j}(1 + \beta_{i+1,j})(1 + \lambda_{i+1,j}) = h_{i+1,j}(1 + \epsilon h_{i+1,j})(1 + \lambda_{i+1,j}) - \hat{h}_i \hat{u}_{ij}.$$  

Therefore,

$$|h_{i+1,j} - \hat{h}_i \hat{u}_{ij} - \hat{u}_{i+1,j}| \leq (2\epsilon + \epsilon^2)[|h_{i+1,j}| + |\hat{u}_{i+1,j}|]$$

and the result follows.
The result obtained in Theorem 2.2 shows that Algorithm 2.1 would be backward stable if \(|\hat{u}_{i,j}| = O(|h_{ij}|)\) for \(j - i \geq 0\). However, it is easy to check that this is not true in general. Just consider the well-known example given below.

\[
H = \begin{bmatrix}
10^{-16} & 1 \\
1 & 1 \\
\end{bmatrix} = \begin{bmatrix}
1 & 0 \\
10^{16} & 1 \\
\end{bmatrix} \begin{bmatrix}
10^{-16} & 1 \\
0 & 1 - 10^{16} \\
\end{bmatrix}.
\]

Hence, the condition for backward stability does not always hold. However, we prove that the LU factorization of upper Hessenberg matrices is stable in the mixed forward-backward sense [5].

**Theorem 2.3.** Let \(H\) be any \(n \times n\) upper Hessenberg matrix with an LU factorization. If \(\hat{L}, \hat{U}\) are the factors obtained by applying Algorithm 2.1 to \(H\), then the following diagram commutes:

\[
\begin{array}{ccc}
H & \xrightarrow{\text{Computed LU}} & \{\hat{L}, \hat{U}\} \\
\downarrow & & \downarrow \\
\hat{H} & \xrightarrow{\text{Exact LU}} & \{\tilde{L}, \tilde{U}\}
\end{array}
\]

where \(\hat{H}\) is obtained from \(H\) by a componentwise relative change not larger than \((2n - 2)\epsilon\), and \(\hat{U}\) is obtained from \(\hat{U}\) by a componentwise relative change no larger than \((2n - 2)\epsilon\).

**Remark 2.4.** In this theorem, \(O(\epsilon^2)\) terms are ignored for simplicity.

**Proof.** As we showed in the proof of Theorem 2.2, the computed quantities satisfy

\[
\hat{l}_i = \frac{h_{i+1,i}(1 + \epsilon_{h_{i+1,i}})(1 + \eta_i)}{\hat{u}_{ii}}, \quad |\epsilon_{h_{i+1,i}}|, |\eta_i| \leq \epsilon,
\]

\[
\hat{u}_{i+1,j}(1 + \beta_{i+1,j})(1 + \lambda_{i+1,j}) = h_{i+1,j}(1 + \epsilon_{h_{i+1,j}})(1 + \lambda_{i+1,j}) - \hat{l}_i \hat{u}_{ij}.
\]

By defining

\[
\tilde{h}_{ij} := h_{ij}(1 + \epsilon_{h_{ij}}) \prod_{k=2}^{i-1} (1 + \beta_{kj}) \prod_{k=2}^{i} (1 + \lambda_{kj}), \quad j \geq i,
\]

\[
\tilde{h}_{i+1,i} := h_{i+1,i}(1 + \epsilon_{h_{i+1,i}})(1 + \eta_i) \prod_{k=2}^{i} (1 + \beta_{ki})(1 + \lambda_{ki}), \quad i = 1, \ldots, n - 1,
\]

\[
\tilde{u}_{ij} := \hat{u}_{ij} \prod_{k=2}^{i} (1 + \beta_{kj})(1 + \lambda_{kj}),
\]

the following exact relations are obtained

\[
\hat{l}_i = \frac{\tilde{h}_{i+1,i}}{\tilde{u}_{ii}}, \quad i = 1, \ldots, n - 1,
\]

\[
\tilde{u}_{i+1,j} = \tilde{h}_{i+1,j} - \hat{l}_i \tilde{u}_{ij}, \quad j = i + 1, \ldots, n.
\]
and the result follows. □

The previous result can be rewritten in the following way:

**Theorem 2.5.** Let $H$ be an $n$-by-$n$ upper Hessenberg matrix with an LU factorization. Let $H = LU$ be the unique $LU$ factorization of $H$ where $L$ is a unit lower triangular matrix. If $\hat{L}$ and $\hat{U}$ are the factors computed by Algorithm 2.1 when applied to $H$, then there exist $\Delta H$ and $\Delta \hat{U}$ such that

$$H + \Delta H = \hat{L}(\hat{U} + \Delta \hat{U}),$$

where $|\Delta H| \leq (2n - 2)\epsilon|H|$ and $|\Delta \hat{U}| \leq (2n - 2)\epsilon|\hat{U}|$.

This means that Algorithm 2.1 is stable in the mixed forward-backward sense.

### 2.1. The LU factorization of Hessenberg matrices is forward stable

According to [5, pag. 10], an algorithm is forward stable if it produces answers with forward errors of similar magnitude to those produced by a backward stable method. This is equivalent to say that an algorithm is forward stable if its forward error divided by the condition number of the problem is $O(\epsilon)$. In this section, we show that Algorithm 2.1 is componentwise forward stable.

The following theorem gives a bound for the forward error obtained when Algorithm 2.1 is applied to a Hessenberg matrix. This bound is given in terms of the condition number of the problem $\text{cond}_{LU}(H)$.

**Theorem 2.6.** Let $H = LU$ be the exact LU factorization of the $n$-by-$n$ upper Hessenberg matrix $H$. Let $\hat{L}$ and $\hat{U}$ be the factors computed by Algorithm 2.1. Then,

$$\max \left\{ \max_{1 \leq i \leq n-1} \left\{ \frac{|l_i - \hat{l}_i|}{|l_i|} \right\}, \max_{1 \leq i \leq n, j \geq i} \left\{ \frac{|u_{ij} - \hat{u}_{ij}|}{|u_{ij}|} \right\} \right\} \leq \frac{(2n - 2)\epsilon}{1 - (2n - 2)\epsilon} \left( 1 + \text{cond}_{LU}(H) \right) + O(\epsilon^2),$$

where $\text{cond}_{LU}(H)$ denotes the condition number of the problem (See Definition 2.7).

**Proof.** By definition of the condition number and taking into account Theorem 2.5, we get

$$\frac{|U - \hat{U} - \Delta \hat{U}|}{|U|} \leq \frac{|\Delta H|}{|H|} \text{cond}_{LU}(H).$$

This implies that

$$\frac{|U - \hat{U}|}{|U|} - \frac{|\Delta \hat{U}|}{|U|} \leq (2n - 2)\epsilon \text{cond}_{LU}(H),$$

or equivalently,

$$\frac{|U - \hat{U}|}{|U|} \leq (2n - 2)\epsilon \text{cond}_{LU}(H) + \frac{|\Delta \hat{U}|}{|U|} \leq (2n - 2)\epsilon \text{cond}_{LU}(H) + (2n - 2)\epsilon \frac{|\hat{U}|}{|U|},$$

Therefore,
\[ \frac{|U - \hat{U}|}{|U|} \leq (2n - 2)\epsilon \text{cond}_{LU}(H) + (2n - 2)\epsilon \left( 1 + \frac{|U - \hat{U}|}{|U|} \right), \]

which proves the result for the entries in \( U \). A bound for the forward errors in the entries in \( L \) can be found in a similar way. Taking into account both bounds, the result follows.

\[ \square \]

From the previous result, taking into account the definition of forward stability, it is enough to prove that \( \text{cond}_{LU}(H) \geq 1 \) to show that Algorithm 2.1 is componentwise forward stable.

Next we give the definition of the relative componentwise condition number of the LU factorization of Hessenberg matrices with respect to relative componentwise perturbations in the entries \( h_{ij} \) of \( H \) with \( i \leq j + 1 \), i.e., \( |\Delta H| \leq u|H| \) with small \( u \).

**Definition 2.7.** Let \( L \) and \( U \) be the matrices obtained from the exact LU factorization of \( H \), where \( H \) is an \( n \times n \) upper Hessenberg matrix. Let \( L + \Delta L \) and \( U + \Delta U \) be the factors obtained from the LU factorization of \( H + \Delta H \). Let us define

\[ DC = \max_{1 \leq j \leq n} \left\{ \frac{|\Delta h_{ij}|}{|h_{ij}|} \right\}, \quad i \leq j + 1, \]

where any quotient has to be understood as zero if the corresponding denominator is equal to zero. Then the relative componentwise condition number of the LU factorization of Hessenberg matrices with respect to small componentwise relative perturbations of \( H \) is defined as

\[ \text{cond}_{LU}(H) := \lim_{u \to 0} \sup_{0 \leq DC \leq u} \max \left\{ \max_{1 \leq i \leq n} \left\{ \frac{|\Delta u_{ij}|}{u_{ij}} \right\}, \max_{1 \leq i \leq (n-1)} \left\{ \frac{|\Delta l_i|}{l_i} \right\} \right\}. \]

The condition number \( \text{cond}_{LU}(H) \) is infinite if some of the denominators appearing in the relative changes of the outputs \( l_i, u_{ij} \), i.e., \( \frac{|\Delta l_i|}{l_i}, \frac{|\Delta u_{ij}|}{u_{ij}} \) is zero. In these cases, other condition numbers have to be considered. For instance, measuring absolute changes in the corresponding components of \( L \) and \( U \), or measuring relative normwise changes of \( L \) and \( U \). We do not consider these particular situations in this work although we mention the second option in the numerical experiments.

The entries of the vector \( l = [l_1, ..., l_{n-1}] \) and the entries of \( U \) in positions \((i, j)\) with \( j \geq i \) are rational functions of the inputs \( h_{ij} \), with \( i \leq j + 1 \), and, as a consequence, the entries of \( U \) and \( l \) are differentiable functions of \( h_{ij} \), whenever the denominators are different from zero. Therefore, \( \text{cond}_{LU}(H) \) can be expressed in terms of partial derivatives [3]. More precisely:
where condLU

Taking into account the definition of Theorem 2.8.

Let

1. This leads to the following result.

$${	ext{max}}_{1 \leq i \leq n, j \geq i} \{\text{condLU}_h(u_{ij})\}, \quad \text{max}_{1 \leq i \leq n-1} \{\text{condLU}_h(l_i)\},$$

condLU

matrix, we performed a number of numerical experiments. In these experiments we

While studying the stability properties of Doolittle’s method for Hessenberg ma-

we repeated the same experiment 20 times in order to consider different random perturbations, and kept the smallest result. In this way, we simulated the forward error we would have got if the algorithm was backward stable. Let us denote

$$\text{condLU}_h(H) = \max_{1 \leq i \leq n, j \geq i} \{\text{condLU}_h(u_{ij})\}, \quad \text{max}_{1 \leq i \leq n-1} \{\text{condLU}_h(l_i)\},$$

where

$$\text{condLU}_h(u_{ij}) = \frac{\sum_{l=0}^{i-1} |h_{il} \frac{\partial u_{ij}}{\partial h_{il}}| + \sum_{k=1}^{i} |h_{kj} \frac{\partial u_{ij}}{\partial h_{kj}}| + \sum_{l=k-1}^{i-1} \sum_{k=2}^{i} |h_{kl} \frac{\partial u_{ij}}{\partial h_{kl}}|}{\text{max}_{1 \leq i \leq n, j \geq i} |u_{ij}|},$$

$$\text{condLU}_h(l_i) = \frac{\sum_{l=0}^{i} |h_{il} \frac{\partial l_i}{\partial h_{il}}| + \sum_{l=k-1}^{i} \sum_{k=2}^{i+1} |h_{kl} \frac{\partial l_i}{\partial h_{kl}}|}{\text{max}_{1 \leq i \leq n-1} |l_i|},$$

Notice that

Taking into account the definition of condLU

Theorem 2.8. Let $H = LU$ be the exact LU factorization of the upper Hessenberg matrix $H$. Let $L$ and $U$ be the factors computed by Algorithm 2.1. Then,

$$\max_{1 \leq i \leq n-1} \left\{ \frac{|l_i - \hat{l}_i|}{|l_i|} \right\}, \quad \max_{1 \leq i \leq n, j \geq i} \left\{ \frac{|u_{ij} - \hat{u}_{ij}|}{|u_{ij}|} \right\} \leq (4n - 4) \text{condLU}_h(H) + O(\epsilon^2),$$

where condLU

3. Numerical experiments

While studying the stability properties of Doolittle’s method for Hessenberg matrices, we performed a number of numerical experiments. In these experiments we compared the forward errors obtained in two different ways:

1) we applied Algorithm 2.1 in standard double precision, i.e., $\epsilon \approx 1.11 \times 10^{-16}$ is the unit roundoff of the finite arithmetic, to 20 random 100-by-100 matrices, and then we compared the results ($\hat{l}_i$ for $i = 1, \ldots, n - 1$; $\hat{u}_i$ for $i = 1, \ldots, n$) with the output of Algorithm 2.1 obtained by using floating point arithmetic with 64 decimal digits of precision ($l_i$ for $i = 1, \ldots, n - 1$; $u_i$ for $i = 1, \ldots, n$). In this way we obtained the “real” forward error, $\text{forw}_r = \max \left\{ \max_{1 \leq i \leq n-1} \left\{ \frac{|l_i - \hat{l}_i|}{|l_i|} \right\}, \max_{1 \leq i \leq n, j \geq i} \left\{ \frac{|u_{ij} - \hat{u}_{ij}|}{|u_{ij}|} \right\} \right\}.$

2) we perturbed randomly each entry $h_{ij}$ of the same random matrices to $\hat{h}_{ij}$ in such a way that $|h_{ij} - \hat{h}_{ij}| \leq 5|h_{ij}|$. Then, we applied Algorithm 2.1 with 64 decimal digits of precision both to the perturbed and nonperturbed inputs, and we compared both outputs. For each random matrix $H$ we repeated the same experiment 20 times in order to consider different random perturbations, and kept the smallest result. In this way, we simulated the forward error we would have got if the algorithm was backward stable. Let us denote
this second forward error as \( forwb \). Namely, if \( lp_{i,k} \) and \( up_{ij,k} \) are the results obtained by applying the algorithm with 64 decimals digits of precision to the \( k \)-th perturbed matrix, then

\[
forwb = \min_k \max_{i,j,k} \left\{ \frac{|l_i - lp_{i,k}|}{|l_i|}, \frac{|u_{ij} - up_{ij,k}|}{|u_{ij}|} \right\}.
\]

The experiments were done using MATLAB 5.3, and we used the variable precision arithmetic of the Symbolic Math Toolbox of MATLAB with 64 decimal digits of precision. In all our tests, theoretical error bounds guarantee that the outputs obtained by running Algorithm 2.1 with 64 decimal digits of precision have more than 50 significant decimal digits.

According to the definition of forward stability, the quotients \( forwr/forwb \) should reflect the forward stability of the algorithm. The same kind of experiments were done with dense 100-by-100 matrices. The only difference is that in step 2) we perturbed randomly each entry \( h_{ij} \) of the random matrices to \( \tilde{h}_{ij} \) in such a way that \( |h_{ij} - \tilde{h}_{ij}| \leq 5 \times 100|\epsilon h_{ij}| \). We recorded for each set of experiments the minimum and the maximum value obtained for the quotient \( forwr/forwb \):

<table>
<thead>
<tr>
<th>( n = 100 )</th>
<th>( \min{forwr/forwb} )</th>
<th>( \max{forwr/forwb} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hessenberg componentwise</td>
<td>3.4 ( 10^{-2} )</td>
<td>( 8 \times 10^{-4} )</td>
</tr>
<tr>
<td>General LU componentwise</td>
<td>0.3</td>
<td>1.7</td>
</tr>
</tbody>
</table>

The results confirm that the LU factorization of Hessenberg matrices is componentwise forward stable. Moreover, these experiments also suggest that the LU factorization of dense matrices should be componentwise forward stable.

Acknowledgment The authors thank Prof. Froilán M. Dopico for his comments and suggestions to improve the contents of this paper.

References


A note on the stability of Hessenberg LU factorization.

C. Brittin
100 Lord Pelham Way
Yorktown, VA 23693
USA
e-mail: cabrittin@gmail.com

M. I. Bueno
Mathematics Department,
The College of William and Mary,
P.O. Box 8795,
Williamsburg, VA 23187-8795 USA
e-mail: mbueno@math.uc3m.es