ABSTRACT

It has been generally assumed that the use of Givens rotations provides significant advantages over the use of Householder transformations for the orthogonal decomposition of sparse matrices. It is also generally acknowledged that the opposite is true for dense matrices. In this paper, a way of applying Householder reflections is described which is competitive with or superior to the use of Givens rotations for sparse orthogonal decomposition. In other words, the advantage of Householder over Givens for dense matrices can carry over to the sparse case, provided that the implementation of the Householder scheme is done in a certain way. The approach relies heavily on the idea of general row merge trees developed by Liu [12]. Results of numerical experiments are provided which demonstrate the advantages of the new scheme. The method also appears to be attractive for use on vector computers.

1. INTRODUCTION

Householder reflections and Givens rotations are standard basic computational operations which are used to compute the orthogonal decomposition of matrices. For a given \( m \) by \( n \) matrix \( A \), a sequence of \( n - 1 \) Householder transformations can be applied to reduce \( A \) to upper triangular form. Alternatively, a sequence of \( m - 1 \) Givens row rotations [that is, a sequence of \( mn - n(n + 1)/2 \) actual rotations] can be used to achieve the reduction.

When the matrix \( A \) is dense, Householder transformations have been employed almost exclusively (LINPACK [2]). However, if the matrix is large and sparse, Givens rotations are generally used or recommended [4, 5, 9]. The scheme proposed by George and Heath [5] makes use of Givens rotations and has the advantage of using a static predetermined data structure for the upper triangular factor matrix. The rows of \( A \) are processed one by one, gradually forming the factor matrix, for which storage has been preallocated. "Intermediate fills" are restricted to the working row, and they are annihilated during the processing of this row.

On the other hand, Householder transformations have been generally regarded as quite inappropriate for sparse QR decomposition [3, 8, 9]. The applications of the Householder column reflections can cause severe intermediate fills. Although they will eventually be annihilated, temporary storage is required to accommodate them, which often turns out to exceed greatly the number of nonzeros in the final factor matrix.

The 8 by 4 matrix example in Figures 1 and 2 serves to illustrate the problem with Householder transformations. The letter "i" is used to denote intermediate fills.

One of the main objectives of this paper is to propose a way of applying Householder transformations so that it becomes competitive with Givens

\[
\begin{array}{cccccccccc}
\times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
\times & & & & & & & & & \\
\times & & & & & & & & & \\
\times & & & & & & & & & \\
\times & \times & & & & & & & & \\
x & x & & & & & & & & \\
x & x & & & & & & & & \\
x & x & & & & & & & & \\
x & x & x & x & x & x & x & x & x & x
\end{array}
\]

![Fig. 1. Sequence of Givens row rotations.](image-url)
rotations for sparse orthogonal decomposition. We hope that this will lead to a reexamination of the role of Householder transformations in sparse computations. Indeed, if this new approach is used, most of the advantages of Householder reflections for the dense case now carry over to sparse systems, while its main disadvantage is removed.

The approach uses the idea of general row merge trees as developed by the second author in [12]. In Section 2, a matrix interpretation of the general row merge scheme is given. We also relate the use of the submatrix annihilation technique to some previous work in the literature.

The basic algorithm is described briefly in Section 3. The main difference between this algorithm and that of [12] is the use of Householder transformations instead of Givens rotations in the core of the numerical QR factorization module. A row-oriented version of Householder reflection is presented to adapt to the computational scheme. A minor modification to the overall merging scheme, motivated by the use of Householder transformations, is also given in that section.

Numerical experiments are provided in Section 4 to compare Householder with Givens. Based on the experimental results reported, Householder reflections do have a role to play in sparse orthogonal factorization. It is consistently faster (in terms of operation counts), and more accurate, in exchange for a very modest increase in working storage. Section 4 also contains our concluding remarks. It is interesting to point out that the row-oriented version of Householder transformations can be adapted to vector computation. Its performance for vector machines, however, will be explored elsewhere.

2. SELECTIVE SUBMATRIX ANNIHILATION

A conventional Givens method [Figure 3(a)] is usually implemented as a row-oriented scheme, in which rows are annihilated one by one using the
partially formed upper triangular factor. On the other hand, the Householder method [Figure 3(b)] is always treated as a column-oriented scheme, and the lower triangular portion of each column is annihilated, column by column.

The proposed scheme in this paper can be viewed as one using a submatrix annihilation technique. Instead of annihilating an entire row or an entire column, a sequence of submatrices within the given sparse matrix is annihilated one after another so that eventually the matrix is reduced to upper triangular.

The added flexibility in the choice of objects to annihilate can lead to major savings in terms of intermediate fills and arithmetic operations. In order to achieve this saving, care must be exercised in the choice of the sequence of submatrices, so that zeros created at one point will not become nonzero again as an intermediate fill at a later stage.

The example in Figure 4 is designed to illustrate possible gains by using submatrix annihilation. The submatrix processed at each step is enclosed by rectangles. Only two intermediate fills (labeled by “i”) occur in this example for this sequence of submatrices. An “f” is used in the figure to represent actual fill in the factor matrix.

FIG. 4. Selective submatrix annihilation sequence.
This new approach of submatrix annihilation originates from the general row merging scheme [12] for the sparse QR decomposition using Givens rotations. It was shown there that the numerical computation in the row merge scheme can be organized as a sequence of reductions of two upper triangular (strictly speaking, trapezoidal) full submatrices into another upper triangular full matrix. Indeed, submatrix annihilation is simply another inter-

![Selective submatrix annihilation for a band matrix.](image)

Fig. 5. Selective submatrix annihilation for a band matrix.
pretation of this scheme, whereby the reduction of two upper triangular submatrices is performed by Householder reflections. For details of the row merge scheme, the reader is referred to [12].

It is interesting to note that this idea of submatrix annihilation has been used implicitly in previous work in the literature. Reid [14] provides an efficient scheme to perform the QR decomposition of a banded system by Householder reductions. It may be interpreted as a wise choice of submatrix annihilation sequence, based on the structure of the band. Figure 5 provides an example.

In [11, Chapter 27], Lawson and Hanson consider an algorithm for the QR decomposition without requiring the entire matrix be in computer storage at one time. That again can be interpreted as a sequence of submatrix annihilations, and in this case the choice depends on the size of the matrix and the amount of available core storage.

3. THE ALGORITHM

3.1 Row Merge Tree

The crucial factors in the proposed approach of submatrix annihilation are the choice of the submatrix sequence and the implementation (or data organization) of the annihilation process.

The notion of row merge trees is introduced in [12]. For an m by n matrix A, a row merge tree is defined to be a strictly binary tree with m leaves, each corresponding to a row in the matrix. It can be used as a basic structure to determine a submatrix sequence, where each submatrix corresponds to a (rooted) subtree. It has the desirable property of preserving zeros created in previous annihilations. Of course, different row merge trees induce different submatrix sequences.

The algorithm proposed here is basically the same as the one in [12], except that Householder reflections are used instead of Givens rotations. The defining row merge trees will be generated by the same heuristic algorithm as described in [12]. In view of the different nature of column reflections and row rotations, there are basic implementational or organizational differences in the process of "merging" or reduction of two upper triangular submatrices. They will be considered in the next subsection.

3.2. Row-Oriented Version of Householder Transformation

Householder transformations have been described and implemented almost exclusively in the form of a sequence of inner products (LINPACK [2]). This is not suitable for our purpose of merging two upper trapezoidal
matrices, since in our scheme, it is more appropriate to store an upper trapezoidal matrix row by row as shown in Figure 6.

In this section, we employ an observation in [10] to show how to reorganize the determination and application of Householder reflections in a row-oriented manner. This new scheme is ideally suited for our computation of reducing two upper trapezoidal matrices stored row by row. It should be noted that it is also suitable for general orthogonal decomposition by Householder transformations on vector machines.

Consider the following $m$ by $n$ matrix:

\[
A = \begin{pmatrix}
d & v^T \\
u & E
\end{pmatrix},
\]

where $u$ and $v$ are $(m - 1)$- and $(n - 1)$-vectors respectively. Let $\sigma$ be the 2-norm of the first column of $A$. That is, let

\[
\sigma = \sqrt{d^2 + u^T u}.
\]

Assume that $\sigma$ is nonzero. Then, the vector $u$ in the matrix $A$ can be annihilated using a Householder reflection given by

\[
P = I - \beta hh^T,
\]

where

\[
h = \begin{pmatrix} 1 \\ z \end{pmatrix},
\]

\[
\beta = 1 + \frac{d}{\sigma_d},
\]

\[
z = \frac{u}{\beta \sigma_d},
\]

\[
\sigma_d = \text{sgn}(d) \sigma.
\]
Here $\text{sgn}(d)$ is a function whose value is $+1$ if $d$ is non-negative, and $-1$ otherwise.

It can be readily verified that

$$h(t^T d) = \sigma_d.$$ 

It is helpful in what follows to define the vectors $q$ and $p$ as

$$q = E^T z \quad \text{and} \quad p = \beta(v + q).$$

**Theorem 3.1.**

$$PA = \begin{pmatrix} d' & v^T \\ 0 & E' \end{pmatrix},$$

where

$$d' = -\sigma_d,$$

$$v^T = v^T - p^T,$$

$$E' = E - zp^T.$$

**Proof.** Consider the application of the Householder transformation to the first column of $A$:

$$P \begin{pmatrix} d \\ u \end{pmatrix} = \begin{pmatrix} d \\ u \end{pmatrix} - \beta hh^T \begin{pmatrix} d \\ u \end{pmatrix}$$

$$= \begin{pmatrix} d \\ u \end{pmatrix} - \beta \sigma_d h = \begin{pmatrix} -\sigma_d \\ 0 \end{pmatrix}.$$ 

On the other hand, applying $P$ to the remaining columns of $A$, we have

$$P \begin{pmatrix} v^T \\ E \end{pmatrix} = \begin{pmatrix} v^T \\ E \end{pmatrix} - \beta hh^T \begin{pmatrix} v^T \\ E \end{pmatrix}$$

$$= \begin{pmatrix} v^T \\ E \end{pmatrix} - \beta h(v^T + z^T E)$$

$$= \begin{pmatrix} v^T \\ E \end{pmatrix} - \begin{pmatrix} 1 \\ z \end{pmatrix} \beta(v^T + q^T)$$

$$= \begin{pmatrix} v^T - p^T \\ E - zp^T \end{pmatrix}.$$
Hence,

\[ E' = E - zp^T \]

and

\[ v'^T = v^T - p^T. \]

The results of Theorem 3.1 suggest the following version of Householder transformations that access the entries of the matrix in a row by row manner.

**Algorithm** (Row-oriented version of Householder transformation).

**Step 1.** Compute \( \sigma_d \) and \( \beta \):

\[
\sigma = \sqrt{d^2 + u^T u} \\
\sigma_d = \text{sgn}(d) \sigma \\
\beta = 1 + \frac{d}{\sigma_d}.
\]

**Step 2.** Compute the factor row \((d', v'^T):\)

(a) \(d' = -\sigma_d\).
(b) Form \(z = u/\beta \sigma_d\).
(c) Compute the appropriate linear combination of the rows of \(A\) (except for the first column) by computing \(q^T = z^T E, \ p^T = \beta(v^T + q^T)\), and

\[ v'^T = v^T - p^T. \]

**Step 3.** Apply Gaussian elimination using the pivot row and column from step 2:

\[
\begin{pmatrix}
1 & p^T \\
z & E
\end{pmatrix}
\rightarrow
\begin{pmatrix}
1 & p^T \\
0 & E'
\end{pmatrix},
\]

where \(E' = E - zp^T\).

Here, we distinguish the *factor row* as the row in the resulting factor matrix \(R\), from the *pivot row* as the row used in the elimination step. It
should be pointed out that one temporary vector is required to store the computed pivot row in step 3.

It is interesting to note that this formulation of Householder transformations can be regarded as a special kind of Gaussian elimination, where the pivot row is computed from a linear combination of the rows of the matrix, rather than being taken directly from it. The numerical stability of the process can be seen from the fact that $|z_i| < 1$ and $|w_i| < 1 < \beta < 2$ for every entry in $z$ and $w$. As usual, care must be exercised in computing $\sigma$ in order to avoid overflows. The standard method involving the scaling of $(d, u)^T$ was employed [2].

The reduction of two upper triangular or trapezoidal submatrices into another upper trapezoidal matrix can now be organized as a sequence of Householder column reflections, where the submatrices are stored in a row by row manner. The basic steps involved can be best illustrated by an example. The following are two such submatrices with column subscript sets \{1,3,4,6,8\} and \{1,3,6,9\} respectively:

\[
\begin{array}{cccccc}
1 & 3 & 4 & 6 & 8 & 1 & 3 & 6 & 9 \\
x & x & x & x & x & x & x & x & x \\
x & x & x & x & x & x & x & x & x \\
x & x & x & x & x & x & x & x & x \\
\end{array}
\]

The merging involves three steps, which are given in Figure 7.

### 3.3. Preliminary Submatrix Reduction

It is well known that the reduction of a matrix with two rows to upper trapezoidal by a Householder reflection is computationally equivalent to the reduction by a Givens rotation [7]. There is, from a practical point of view, no advantage of Householder over Givens in such situations.

On the other hand, Householder reflections will be more effective than Givens rotations if the matrix to be reduced has many rows. In this case, one reflection can be used instead of many rotations to reduce all nonzeros under the diagonal in the first column. In view of this, the row merging sequence given in [12] is modified so that the algorithm will accumulate as many rows as possible before reduction by Householder reflections is performed. The criterion used is that the algorithm will take the next incoming row if it does not enlarge the subscript set.

In Figure 8, a row merge tree is specified for the given 8 by 5 matrix example. This means that seven reductions of submatrices are to be performed. Since the first four rows have the same set of nonzero subscripts, it will be more advantageous to reduce them together by Householder transformations. The same applies to the last four rows. The row merge tree structure can hence be depicted as in Figure 9. Note that if one is using Givens
Step 1. Union of column subscripts:

\[ 1 \ 3 \ 4 \ 6 \ 8 \ 9 \]

Step 2. Extension of the submatrices:

\[
\begin{align*}
&1 \ 3 \ 4 \ 6 \ 8 \ 9 \\
&x \ x \ x \ x \ x \\
&x \ x \ x \\
&x \ x \\
&x & x \\
&x & x \\
&x & x \\
&x & x
\end{align*}
\]

Step 3. Sequence of Householder reflections (see Figure 7).

Fig. 7. Merging two submatrices.

rotations, the two row merge structures are equivalent from a computational point of view.

4. EXPERIMENTAL RESULTS AND CONCLUDING REMARKS

In this section, experimental results for sparse orthogonal decomposition using Householder reflections and Givens rotations are provided. The times reported in the tables are in seconds on a VAX 11/780 having floating point
Our experiments involved two sets of test problems which display a considerable variety of structures. For test set 1, the matrix values were obtained directly from the application, while for test set 2, the numerical values of the nonzeros in the matrices were uniform random numbers from [-1, 1]. In all cases, the solution $x$ was arranged to be a vector of all ones by setting the right hand side $b$ equal to the sum of the columns of the matrix, computed in double precision. In the tables, the relative error reported is the maximum relative error observed over all components in $x$, and the residual reported is the maximum value of $|r_i|$, taken over all $i$, where $r = b - Ax$, computed in double precision. All other computations were done in single precision.

The first test set consists of the ten problems used by George, Heath, and Ng in their comparison paper on methods for solving sparse linear least
TABLE 1
MATRIX PROBLEMS FOR TEST SET 1

<table>
<thead>
<tr>
<th>Problem</th>
<th>Number of</th>
<th>Problem description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>313 176 1557</td>
<td>Sudan survey data</td>
</tr>
<tr>
<td>2</td>
<td>1033 320 4732</td>
<td>Analysis of gravity-meter observations (well-conditioned)</td>
</tr>
<tr>
<td>3</td>
<td>1033 320 4719</td>
<td>Analysis of gravity-meter observations (ill-conditioned)</td>
</tr>
<tr>
<td>4</td>
<td>1850 712 8755</td>
<td>Similar to Problem 2, but larger</td>
</tr>
<tr>
<td>5</td>
<td>1850 712 8638</td>
<td>Similar to Problem 4, but larger</td>
</tr>
<tr>
<td>6</td>
<td>784 225 3136</td>
<td>15 x 15 grid problem</td>
</tr>
<tr>
<td>7</td>
<td>1444 400 5776</td>
<td>20 x 20 grid problem</td>
</tr>
<tr>
<td>8</td>
<td>1512 402 7152</td>
<td>3 x 3 geodetic network with 2 observations per node</td>
</tr>
<tr>
<td>9</td>
<td>1488 784 7040</td>
<td>4 x 4 geodetic network with 1 observation per node</td>
</tr>
<tr>
<td>10</td>
<td>900 269 4208</td>
<td>Geodetic network problem provided by U.S. National Geodetic Survey (ill-conditioned)</td>
</tr>
</tbody>
</table>

Readers are referred to it for details about the problems. We list them in Table 1 for reference.

Various performance results of the different methods applied to test set 1 are tabulated in Tables 2 and 3. The column ordering used was that provided by the minimum degree algorithm, as suggested in [5]. A modified form of the minimum degree algorithm due to Liu [13] was actually used in the experiments.

The second set of test problems are typical of those that arise in the natural factor formulation of finite element methods [11]. Consider the k by k regular grid with \((k - 1)^2\) small squares. Associated with each of the \(k^2\) grid nodes is a variable, and associated with each square is a set of four equations involving the four variables at the corners of the square. The assembly of these equations results in a large sparse overdetermined system of equations

\[
Ax = b,
\]

where the matrix A has \(k^2\) columns and \(4(k - 1)^2\) rows. The columns of the matrix were ordered by the minimum degree scheme as in [5]. The different schemes were tested on values of \(k = 10, 20, 30, 40,\) and 50, and the results are tabulated in Tables 4 and 5.
TABLE 2

COMPARISON OF FACTORIZATION OPERATION COUNTS AND TIMES FOR THE DIFFERENT SCHEMES APPLIED TO THE PROBLEMS IN TEST SET 1

<table>
<thead>
<tr>
<th>Problem</th>
<th>Factorization opcount</th>
<th>Factorization time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preproc Givens</td>
<td>Hholder</td>
</tr>
<tr>
<td>1</td>
<td>51,732</td>
<td>51,306</td>
</tr>
<tr>
<td>2</td>
<td>141,744</td>
<td>149,068</td>
</tr>
<tr>
<td>3</td>
<td>143,764</td>
<td>149,049</td>
</tr>
<tr>
<td>5</td>
<td>477,920</td>
<td>445,960</td>
</tr>
<tr>
<td>6</td>
<td>138,320</td>
<td>120,002</td>
</tr>
<tr>
<td>7</td>
<td>357,444</td>
<td>285,182</td>
</tr>
<tr>
<td>8</td>
<td>330,884</td>
<td>333,970</td>
</tr>
<tr>
<td>9</td>
<td>382,936</td>
<td>365,286</td>
</tr>
<tr>
<td>10</td>
<td>561,156</td>
<td>465,598</td>
</tr>
</tbody>
</table>

TABLE 3

COMPARISON OF NUMERICAL RESULTS FOR THE DIFFERENT METHODS APPLIED TO THE PROBLEMS IN TEST SET 1

<table>
<thead>
<tr>
<th>Problem</th>
<th>Relative error</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preproc Givens</td>
<td>Hholder</td>
</tr>
<tr>
<td>1</td>
<td>2.86e-6</td>
<td>1.67e-6</td>
</tr>
<tr>
<td>2</td>
<td>0.89e-5</td>
<td>0.68e-5</td>
</tr>
<tr>
<td>3</td>
<td>0.35e-3</td>
<td>0.27e-3</td>
</tr>
<tr>
<td>4</td>
<td>3.81e-6</td>
<td>3.52e-6</td>
</tr>
<tr>
<td>5</td>
<td>1.04e-4</td>
<td>0.99e-4</td>
</tr>
<tr>
<td>6</td>
<td>1.25e-6</td>
<td>1.19e-6</td>
</tr>
<tr>
<td>7</td>
<td>1.31e-6</td>
<td>1.25e-6</td>
</tr>
<tr>
<td>8</td>
<td>1.79e-6</td>
<td>1.79e-6</td>
</tr>
<tr>
<td>9</td>
<td>2.15e-6</td>
<td>3.46e-6</td>
</tr>
<tr>
<td>10</td>
<td>3.87e-3</td>
<td>8.18e-3</td>
</tr>
</tbody>
</table>

The results of the numerical experiments demonstrate that the use of Householder reflections in sparse matrix decomposition can be organized so that they are very competitive with Givens rotations. For the variety of problems in test set 1, their overall performance is at least as good as that of Givens. The factorization operation count is always smaller, while the actual CPU time is always comparable.
For test set 2, there is a more substantial reduction in the operation count. Indeed, for $k = 50$, the operation count was reduced by a factor of more than one third. It should be noted that the reduction in arithmetic is not reflected in a proportional decrease in execution time. This is probably because of larger computational overhead incurred in the Householder version. Nevertheless, for $k = 50$, the preprocess version of the new scheme yields a reduction in execution time of nearly 25 percent.

In summary, the experiments confirm that the widely held view that Householder transformations are inappropriate for sparse orthogonal decomposition should be reassessed. In particular, we have shown here that the preprocess version is competitive with and often superior to the use of Givens rotations.

It would appear that the row-oriented version of Householder reflections described in this paper is readily adaptable to vectorization. We are currently
attempting to conduct such experiments on a Cray computer, and we will report the results of those experiments when they have been completed.

The authors are grateful to the referee for making some helpful suggestions which significantly improved the presentation in Section 3.

REFERENCES


Received 26 November 1984; revised 24 June 1985