The Computation of Elementary Unitary Matrices

R.B. Lehoucq*

August 28, 1995

Abstract

The construction of elementary unitary matrices that transform a complex vector to a multiple of e_1, the first column of the identity matrix, are studied. We present four variants and their software implementation, including a discussion on the LAPACK subroutine CLARFG. Comparisons are also given.

1 Introduction

The goal of this paper is to survey elementary unitary matrices. We begin by first discussing elementary unitary matrices that are Hermitian. Let w be a complex vector. Define the elementary Hermitian matrix $U = I - 2ww^H$, where $w^H w = 1$. It is easily verified that U is both Hermitian and unitary. In particular, if w is a real vector, then U is orthogonal and symmetric, and is commonly referred to as a Householder reflector. Since U is unitary, its inverse is readily available.

Two important applications of elementary Hermitians include the computation of the QR factorization of a matrix, and the orthogonal reduction of a square matrix A into upper Hessenberg form. The former application is often used for the stable computation of a solution for the linear least squares problem. The latter application is needed for many eigenvalue computations. The literature on elementary Hermitians is vast. For information on applications concerning Householder matrices see Golub and Van Loan [4]. Parlett [7] examines the algorithmic and stability issues of computing Householder matrices. A detailed error analysis by Wilkinson [10] shows the stability of numerical techniques using elementary Hermitians. Besides these excellent numerical properties, their application demonstrates their efficiency. If A is a matrix, then $UA = A - 2w(A^H w)^H$, and hence explicit formation and storage of U is not required. Only the ability to form the matrix-vector product $A^H w$ and a rank one update to A.

Fundamental to the use of elementary Hermitians in the above applications is their ability to transform a vector x to a multiple of e_1, the first column of the identity matrix. As we will show, an elementary Hermitian is not always defined when x is to be transformed to a real multiple of e_1. However, the crucial property of unitariness may be preserved. The purpose of this paper is to review and examine the details of constructing an elementary unitary matrix so that a complex vector x is transformed to a multiple of e_1.

*Department of Computational and Applied Mathematics, Rice University. (lehoucq@rice.edu)
This work was supported by DARPA under contract TV-ORA4466.01

The paper is organized as follows. In § 2 the mathematical problem is stated and general conditions for constructing elementary unitary matrices are derived. The four approaches for construction are then introduced in § 2.1–§ 2.4. The first one is implemented in EISPACK [8] and is based upon a development by Wilkinson [9, pages 48–50]. The LINPACK [2] approach is the second one studied. The third approach is due to Hammarling and Du Croz. It is implemented in the NAG Fortran Library subroutine F06HRF [6]. The final variation is implemented by the LAPACK [1] subroutine CLARFG. The details of this software implementation are also discussed. Section three is a comparison and summary of our findings. In fact, our attempt to understand the differences between the Wilkinson approach and the alternate formulation implemented by LAPACK led to this study.

We employ Householder notational conventions. Capital and lower case letters denote matrices and vectors, respectively, while lower case Greek letters denote scalars. In particular, $\xi_i = e_i^T x$ denotes the i-th element of the vector x. Unless otherwise stated, all quantities are assumed to be complex and $i \equiv \sqrt{-1}$. The real and imaginary part of a complex number α are denoted by $Re(\alpha)$ and $Im(\alpha)$, respectively. The vector norm used is the Euclidean one: $\|x\| = \sqrt{x^H x}$. The reader is also reminded that $|\alpha|^2 = \bar{\alpha}\alpha$ where $\bar{\alpha}$ is the complex conjugate of α.

2 Elementary Unitary matrices

Let us clearly state the problem at hand. Find an elementary unitary matrix that satisfies the following three conditions:

$$U = I - \sigma w w^H, \quad U^H x = \gamma \|x\| e_1, \quad |\gamma| = 1, \tag{1}$$

where x is a vector with n components. The third condition is a consequence of the second one since $\|U^H x\|/\|x\| = |\gamma|$. The second condition gives that $x^H U^H x = \gamma \|x\| x^H e_1$ implying that U is an elementary Hermitian matrix if and only if σ and $\gamma x^H e_1$ are real.

The matrix U as defined by (1) is a special member of the more general class of *elementary* matrices defined by

$$E(w, v; \sigma) \quad = \quad I - \sigma w v^H. \tag{2}$$

See Householder [5] and Wilkinson [11] for introductions. Dubrulle [3] presents a comprehensive study for the case of real w,v and σ, that includes a discussion to *block* implementations.

Let us determine general conditions for an elementary matrix to be unitary. Since $E(w, v; \sigma)$ must be unitary,

$$I = (I - \sigma w v^H)^H (I - \sigma w v^H) = I - \bar{\sigma} v w^H - \sigma w v^H + \sigma\bar{\sigma}(w^H w) v v^H.$$

Cancelling terms results in

$$\sigma\bar{\sigma}(w^H w) v v^H \quad = \quad \bar{\sigma} v w^H + \sigma w v^H. \tag{3}$$

Rearranging terms gives $(\sigma\bar{\sigma}(w^H w) v - \sigma w) v^H = \bar{\sigma} v w^H$, and a row space argument implies that w and v are linearly dependent. Substituting $v = w$ into (3) gives

$$|\sigma|^2 \|w\|^2 = \sigma + \bar{\sigma} = 2Re(\sigma) \tag{4}$$

as the required relationship between σ and w. Note that the above relationship contains some redundancy. Scaling w by a complex number η and dividing σ by $|\eta|^2$ still satisfy the relationship. This scaling also satisfies the second condition of (1) since $(\bar{\sigma}|\eta|^{-2})(w\eta)(w\eta)^H = \sigma w w^H$. Finally, the second condition of (1) gives that w is a linear combination of x and e_1.

Four sets choices for w, σ and γ are the subject of the § 2.1–2.2. A standard modification for $w = \mu x + \nu e_1$ is that $\mu\xi_1$ and ν share the same sign. In floating point arithmetic, this choice of sign leads to a small relative error when computing w. For example, if $\mu = 1$ the sign of e_1 is that of $Re(\xi_1)$. Parlett [7] presents a thorough discussion on the choice of sign when computing Householder reflectors. For the remainder of the paper, $\nu \equiv \mathrm{Sign}(Re(\xi_1))\|x\|$.

Note that an elementary Hermitian (and Householder) matrix chooses $w = (x + \nu e_1)/\|x + \nu e_1\|$ so that $w^H w = 1$, $\gamma = -1$. Conditions (1) and (4) are satisfied.

2.1 The Wilkinson Approach

Wilkinson [9, pages 49–50] suggested the following modification. Let $\xi_1 = e^{i\theta_1}|\xi_1|$ where $0 \le \theta_1 < 2\pi$ and

$$x = e^{i\theta_1} y = e^{i\theta_1}[|\xi_1|, e^{-i\theta_1}\xi_2, \ldots, e^{-i\theta_1}\xi_n]^T.$$

Then even if ξ_1 has a non-zero imaginary part, $e_1^T y$ is a real number, an elementary Hermitian P may be constructed so that Py is a real multiple of e_1. Thus, condition (4) is satisfied as already discussed. Set $U = e^{i\theta_1}P$ and

$$U^H x = (e^{-i\theta_1}P)(e^{i\theta_1}y) = Py = \gamma\|x\|e_1,$$

where $\gamma = -1$. The matrix U is a multiple of an elementary unitary matrix. Since the first component of y is a non-negative number, θ_1 is zero.

Although EISPACK [8] does not have a subroutine that computes an elementary unitary matrix, the subroutines CORTH and HTRIDI implement a slight variation of the Wilkinson approach. CORTH [8, pages 300–305] and HTRIDI [8, pages 357–363] orthogonally reduce a general and Hermitian matrix to upper Hessenberg and tridiagonal form, respectively. They set $U = P$ directly and thus transform y to $-e^{i\theta_1}\|x\|e_1$. The software sets $w = x + e^{i\theta_1}\|x\|e_1 (= e^{i\theta_1}(y + \|x\|e_1))$ and $\sigma = 2(w^H w)^{-1}$. Hence $w^H w = 2\|x\|(\|x\| + |\xi_1|)$ and $\sigma = 1/\|x\|(\|x\| + |\xi_1|)$ thus satisfying condition (4). A simple calculation shows that

$$U^H x = x - \bar{\sigma}(w^H x)x = x - \sigma\|x\|(\|x\| + |\xi_1|)x = \gamma\|x\|e_1,$$

where $\gamma = -e^{i\theta_1}$. In order to prevent possible overflow when computing σ, the vector x is is initially normalized by $\theta = |Re(\xi_1)| + |Im(\xi_1)| + \cdots + |Re(\xi_n)| + |Im(\xi_n)|$.

2.2 The LINPACK Approach

As in EISPACK, LINPACK does not have a general purpose subroutine implementing the solution of problem (1). However, subroutines CQRDC [2, chapter 9] and CSVDC [2, chapter 11] employ elementary unitary matrices. Subroutines CQRDC and CSVDC compute the QR factorization and singular value decomposition of a complex matrix, respectively.

The LINPACK form for an elementary unitary matrix is easily derived by scaling the w used by EISPACK with $\eta = e^{-i\theta_1}/\|x\|$. From the remarks regarding the scaling of equation (4), $\sigma = \|x\|/(\|x\| + |\xi_1|)$ and the LINPACK U is such that $U^H x = \gamma \|x\| e_1$ where $\gamma = -e^{i\theta_1}$. Note that for non-zero x, $.5 \leq \sigma \leq 1$ thus avoiding the risk of overflow possible in the in the (unscaled) EISPACK variant.

2.3 The NAG Approach

The second form for an elementary unitary matrix is due to Hammarling and Du Croz [6], (Introduction – F06). Unlike the previous two versions, this one computes an elementary unitary matrix U so that $U^H x$ is a *real* multiple of e_1. As explained at the beginning of § 2, the resulting σ cannot be real unless ξ_1 is also.

Choosing $\sigma = (x^H w)^{-1}$ where $w = x + \nu e_1$ results in $U^H x = (I - \bar{\sigma} w w^H) x = x - (\bar{\sigma} w^H x) w = \gamma \nu e_1$ where $\gamma = -1$. This choice of σ will satisfy (4) as we now demonstrate. First

$$w^H x = (x^H + \nu e_1^T) x = x^H x + \nu \xi_1 = \nu(\nu + \xi_1),$$

which determines σ and $\|w\|^2 = (x^H + \nu e_1^T)(x + \nu e_1) = 2\nu(\nu + \text{Re}(\xi_1))$. Finally

$$\begin{aligned} (w^H x)(x^H w)(\sigma + \bar{\sigma}) &= (w^H x)(x^H w)(\frac{1}{w^H x} + \frac{1}{x^H w}), \\ &= x^H w + w^H x, \\ &= \nu(\nu + \bar{\xi_1}) + \nu(\nu + \xi_1), \\ &= 2\nu(\nu + \text{Re}(\xi_1)), \end{aligned}$$

shows that $|\sigma|^2(\sigma + \bar{\sigma}) = \|w\|^2$ as claimed. Note that when ξ_1 is real, U is Hermitian. This version does not appear to be as widely known as the Wilkinson one.

The NAG subroutine F06HRF computes an elementary unitary matrix so that

$$Re(|\eta|^{-2}\sigma) = 1 \quad \text{and} \quad 1 \leq e_1^T \eta w \leq \sqrt{2}, \tag{5}$$

for some scale factor η. First note that $e_1^T w/(\xi_1 + \nu) = 1$. Then, from the manner in which ν was chosen, it follows that $Re(\sigma|\xi_1 + \nu|^{-2}) = (\|x\| + |Re(\xi_1)|)/\|x\|$. Hence the choice of $\eta = \sqrt{(\|x\| + |Re(\xi_1)|)/\|x\|}/(\xi_1 + \nu)$ is such that

$$e_1^T \eta w = \sqrt{\frac{\|x\| + |Re(\xi_1)|}{\|x\|}} \quad \text{and} \quad |\eta|^{-2}\sigma = \frac{\|x\| + \text{Sign}(Re(\xi_1))\xi_1}{\|x\| + |Re(\xi_1)|},$$

and the two conditions (5) on η are satisfied. Note that $1 \leq |\eta|^{-2}|\sigma| \leq 2$.

2.4 The LAPACK approach

The LAPACK subroutine CLARFG is a slight variant of the one used by the NAG subroutine F06HRF. The resulting code is an excellent example of the art of developing software from a numerical algorithm. Using the notation of the previous section for w and σ, let $\eta^{-1} = \xi_1 + \nu$ and hence $e_1^T \eta w = 1$ and $|\eta|^{-2}\sigma = (\xi_1 + \nu)/\nu$. Conditions (1)

Problem Statement:

Compute $U = I - \sigma w w^H$ where $U^H x = \gamma \|x\| e_1$, $U^H U = I$, and $|\gamma| = 1$.

Notation:

$\xi_i = e_i^T x$ for $i = 1:n$, $\quad \nu = \text{Sign}(Re(\xi_1))\|x\|$, $\xi_1 = e^{i\theta_1}|\xi_1|$ where $0 \le \theta_1 < 2\pi$, $\quad \kappa = (|Re(\xi_1)| + \|x\|)/\|x\|$			
Method	w	σ	γ
EISPACK	$x + e^{i\theta_1}\|x\|e_1$	$1/\|x\|(\|\xi_1\| + \|x\|)$	$-e^{i\theta_1}$
LINPACK	$xe^{-i\theta_1}/\|x\| + e_1$	$\|x\|/(\|\xi_1\| + \|x\|)$	$-e^{i\theta_1}$
NAG	$(x + \nu e_1)\sqrt{\kappa}/(\xi_1 + \nu)$	$(\xi_1 + \nu)/\nu\kappa$	-1
LAPACK	$(x + \nu e_1)/(\xi_1 + \nu)$	$(\xi_1 + \nu)/\nu$	-1

Table 1: Comparisons for the four variants used to compute an elementary unitary matrix

and (4) are satisfied since w and σ are scaled here by η and $|\eta|^{-2}$, respectively. Note that $1 \le |\eta|^{-2}|\sigma| \le 2$. If x is a real multiple of e_1 then $\tau \leftarrow 0$ and $U \leftarrow I$.

Representing U for use in further computation only requires storage for the complex τ. The storage for x may be re-used to write both ν and the *essential* part of w, that is $x \leftarrow [\nu, \xi_2/(\xi_1 + \nu), \ldots, \xi_n/(\xi_1 + \nu)]^T$.

One who reviews subroutine CLARFG will notice the programmer took care not to reciprocate the number $\|x\|$ that may fall below a certain machine dependent tolerance, SAFMIN. The value SAFMIN, computed by the LAPACK auxiliary subroutine SLAMCH is a machine dependent lower bound for numbers that may be safely reciprocated and not cause an overflow condition. If $\|x\|$ is less than the lower bound then the vector x is scaled by a multiple of the reciprocal of SAFMIN until it is at least as large as SAFMIN. Defining the integer k to represent the number of scalings required, let $\theta = k/\text{SAFMIN}$. The number σ may now be safely computed as $\sigma \leftarrow (\nu + \theta\xi_1)/\nu$ where $\nu \leftarrow \text{Sign}(Re(\theta\xi_1))(\|\theta x\|)$. The essential part of u is computed as $(\theta\xi_1 + \theta\nu)^{-1}[\theta\xi_2, \ldots, \theta\xi_n]^T$. This same scaling technique is also used by the real precision version of CLARFG—SLARFG.

3 Comparisons and Conclusions

Four different forms of elementary unitary matrices were presented to solve the elimination problem defined by (1). Table 1 presents a summary of the four approaches. We now briefly analyze the information in the table.

- The EISPACK approach. Benefit: Real σ. Cost: An initial scaling of x to prevent possible overflow when computing σ and storing a possibly complex γ.
- The LINPACK approach. Benefit: Real σ; $.5 \le |\sigma| \le 1$. Cost: Storing a possibly complex γ.

- The NAG approach. Benefit: Directly obtains a real γ. Cost: Storing a possibly complex σ and forming a square root; $1 \leq |\sigma| \leq 2$.

- The LAPACK approach. Benefit: Directly obtains a real γ. Cost: Storing a possibly complex σ; $1 \leq |\sigma| \leq 2$.

Examining the application of U to a matrix A allows the following analysis:

- The EISPACK and LINPACK approaches require computing $A - \sigma w(A^H w)^H$ with real σ.

- The LAPACK and NAG compute $A - \sigma w(A^H w)^H$ with possibly complex σ.

Since computing the QR factorization of a matrix,the bidiagonal, Hessenberg, and tridiagonal reductions, involve applications of elementary unitary matrices to A, the computation is always cheaper with real σ.

The benefit of directly computing a real γ is that it allows reuse of software. For example, when reducing a Hermitian matrix to tridiagonal form, the resulting tridiagonal matrix is real, and the symmetric tridiagonal QR algorithm may then be employed [1]. The same may be said about the preliminary reduction of a matrix to bidiagonal form needed by the singular value decomposition: see [1, page 42] and [2, chapter 9]. A third example is when computing a QR factorization of a matrix A. For stable computation of a solution to a linear least squares problem, a triangular system of equations involving R is often required. Directly computing a real γ results in real numbers on the diagonal of R. Thus the careful scaling algorithms used by LAPACK when solving triangular system of equations may be employed.

On the other hand, when using either the EISPACK and LINPACK forms of elementary unitary matrices, a diagonal unitary matrix D may always be computed to allow reuse of software or the use of careful scaling algorithms. For example, when computing a QR factorization of a matrix A with m rows and n columns, let $D = \mathrm{Diag}(\delta_1, \ldots, \delta_m)$ be the diagonal matrix where $\delta_j = e_j^T R e_j / |e_j^T R e_j|$ for $j = 1 : \min(m, n)$ and $\delta_j = 1$ otherwise. It then follows that $A = QR = (QD)(D^H R)$, QD is unitary and the diagonal elements of $D^H R$ are real numbers. Similar procedures may be employed when further reducing a Hermitian tridiagonal matrix to real symmetric tridiagonal form and when reducing a matrix to real bidiagonal form. Further computation and storage is required. The elementary unitary matrices based on the Hammarling and Du Croz approach implicitly perform this post-processing step.

4 Acknowledgments

The author would like to thank Jeremy Du Croz and Dan Sorensen for background information on the Hammarling–Du Croz approach and for encouragement. An anonymous referee, the handling editor W. Van Snyder and Leslea Davison made many helpful remarks that improved the initial submission of the manuscript. The clever scaling in § 2.4 is due to James Demmel.

References

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK Users' Guide*. SIAM, Philadelphia, PA., second edition, 1992.

[2] J.J. Dongarra, C.B. Moler, J.R. Bunch, and G.W. Stewart. *LINPACK Users' Guide*. SIAM, Philadelphia, PA., 1979.

[3] A. A. Dubrulle. Work notes on elementary matrices. Technical Report HPL-93-69, Hewlett–Packard Laboratories, 1993.

[4] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins, second edition, 1989.

[5] A. S. Householder. *The Theory of Matrices in Numerical Analysis*. Dover, 1974.

[6] Numerical Algorithms Group Ltd. *NAG Fortran Library Manual, Mark 16*. Oxford, 1993.

[7] B. N. Parlett. Analysis of algorithms for reflectors in bisectors. *SIAM Review*, 13(2):197–208, April 1971.

[8] B. T. Simth, J. M. Boyle, J. J. Dongarra B. S. Garbow, Y. Ikebe, V. C. Klema, and C. B. Moler. *EISPACK Guide*. Springer–Verlag, Berlin, second edition, 1976. Volume 6 of Lecture Notes in Computer Science.

[9] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, UK, 1965.

[10] J. H. Wilkinson. Error analysis of transformatins based on the use of matrices of the form $I - 2ww^H$. In L.B. Rall, editor, *Error in Digital Computation*, volume 2, pages 77–101. John Wiley, 1965.

[11] J. H. Wilkinson. Some recent advances in numerical linear algebra. In D.A.H. Jacobs, editor, *The state of the Art in Numerical Analysis*, pages 1–53. Academic Press, 1977.