A Basis-Kernel Representation of Orthogonal Matrices*

Xiaobai Sun and Christian Bischof

Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439
{xiaobai,bischof}@mcs.anl.gov

Argonne Preprint MCS-P431-0594

Abstract. In this paper we introduce a new representation of orthogonal matrices. We show that any orthogonal matrix can be represented in the form $Q = I - YSY^T$, which we call the basis-kernel representation of Q. We show that the kernel S can be chosen to be triangular and show how the familiar representation of an orthogonal matrix as a product of Householder matrices can be directly derived from a representation with triangular kernel. We also show that there exists an, in some sense, minimal orthogonal transformation for solving the block elimination problem. We explore how the basis Y determines the subspaces that Q acts on in a nontrivial fashion, and how S determines the way Q acts on this subspace. We derive a canonical representation that explicitly shows how Q partitions $\mathbf{R}^n$ into three invariant subspaces in which it acts as the identity, a reflector, and a rotator, respectively. We also derive a generalized Cayley representation for arbitrary orthogonal matrices, which illuminates the degrees of freedom we have in choosing orthogonal matrices acting on a predetermined subspace.

Key words. Orthogonal Matrices, Block Elimination, Orthogonality Condition, Basis-Kernel Representation, Cayley Transform, Householder Matrices.

1 Introduction

Orthogonal transformations are a well-known tool in numerical linear algebra and are used extensively in decompositions such as the QR factorization, tridiagonalization, bidiagonalization, Hessenberg reduction, or the eigenvalue or singular value decomposition of a matrix (see, for example, [6,

---

*This work was supported by the Applied and Computational Mathematics Program, Advanced Research Projects Agency, under contract DM28E04120, and by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38. This paper is PRISM Working Note #19, available via anonymous ftp to `ftp.super.org` in the directory pub/prism.

10]). The orthogonal transformations employed are usually compositions of the following elementary transformations:

Givens Rotator:

$$G = G(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \tag{1}$$

In the two-dimensional plane, application of G to a vector x amounts to a clockwise rotation of x by an angle of θ.

Jacobi Reflector:

$$J = J(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix} \tag{2}$$

In the two-dimensional plane, application of J to a vector x amounts to reflecting x with respect to the line spanned by the vector

$$(\cos(\theta/2), \sin(\theta/2))^{\mathrm{T}}.$$

Householder Reflector:

$$H = H(v) = I - \beta v v^{\mathrm{T}}, \qquad \beta v^{\mathrm{T}} v \beta = 2\beta. \tag{3}$$

This representation of Householder matrices is used in the LINPACK [4] and LAPACK [1] libraries. The condition on v and β in (3) covers all choices for v and β that result in an orthogonal matrix H. In particular, it includes the degenerate case $\beta = 0$ where H is the identity matrix I. Note that the application of H to a vector x amounts to a reflection of x with respect to the hyperplane $\mathcal{R}(v)^{\perp}$, the orthogonal complement of the range $\mathcal{R}(v)$.

Each of the three well-known elementary transformations, when applied to a matrix, implies a low-rank (rank 1 or 2) update of the matrix.

Givens rotators form a group under matrix multiplication with the identity matrix as the unit element of the group; in particular, the product of any two Givens rotators is again a Givens rotator. Note that unless $\theta = 0 \bmod 2\pi$, $G(\theta)$ has no eigenvalue at 1. That is, except for the identity, a Givens reflector rotates every nonzero vector in the entire two-dimensional space.

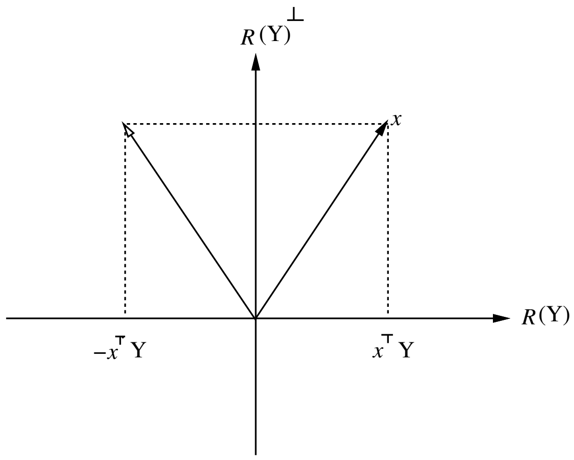

Figure 1: Reflectors

In contrast, Jacobi reflectors are not closed under matrix multiplication. As a matter of fact, the product of any two reflectors is a rotator. A Jacobi reflector can be represented as a rank-1 modification to the identity matrix, namely,

$$J(\theta) = I - (I - J) = I - 2yy^{\mathrm{T}}, \quad \text{where } y = \begin{pmatrix} \sin(\theta/2) \\ -\cos(\theta/2) \end{pmatrix}. \tag{4}$$

Unlike Givens rotation, a Jacobi reflector divides $\mathbf{R}^2$ into two complementary subspaces, acting as the identity on one of them and reflecting on the other:

$$Jx = \begin{cases} x & x \in \mathcal{R}(y)^{\perp}, \\ -x & x \in \mathcal{R}(y). \end{cases}$$

For an arbitrary vector $x \in \mathcal{R}^2$, $J(\theta)x$ is therefore a reflection of x with respect to the line $\mathcal{R}(y)^{\perp} = \mathcal{R}([\cos(\theta/2), \sin(\theta/2)]^{\mathrm{T}})$. We may also say Jx is the reflection of x along $\mathcal{R}(y)$, or simply along y. For the special Jacobi reflector $J(0)$, $J(0) = J(2\pi) = I - 2e_2e_2^{\mathrm{T}}$. This is illustrated in Figure 1.

A Givens rotator $G(\theta)$ can always be represented as a product of two Jacobi reflectors,

$$G(\theta) = J(\alpha)J(\beta), \quad \text{with} \quad \beta - \alpha = \theta \bmod 2\pi.$$

In particular, $G(\theta) = J(0)J(\theta)$. That is, $G(\theta)$ can be decomposed as a reflection with respect to $(\cos(\theta/2), \sin(\theta/2))^{\mathrm{T}}$ followed by another reflection with respect to $(1, 0)^{\mathrm{T}}$. Thus $G(\theta)$ can be represented as a rank-2 modification to the identity matrix,

$$G(\theta) = I - YSY^{\mathrm{T}}, \tag{5}$$

with, for instance,

$$Y = \begin{pmatrix} 0 & \sin(\theta/2) \\ 1 & -\cos(\theta/2) \end{pmatrix}, \text{ and } S = \begin{pmatrix} 2 & 4\cos(\theta/2) \\ & 2 \end{pmatrix}.$$

Householder reflectors are a direct generalization of Jacobi reflectors. For each vector x, $H(v)x$ is the reflection of x with respect to the hyperplane $\mathcal{R}(v)^{\perp}$. The concept of reflectors was further developed by Schreiber and Parlett [8] to *block reflectors*

$$Q = I - 2YY^{\mathrm{T}}, \qquad Y^{\mathrm{T}}Y = I, \quad Y \in \mathcal{R}^{m\times k}. \tag{6}$$

Note that the reflectors we have mentioned so far are all symmetric.

The representations (3), (4), and (6) for reflectors and (5) for rotators are all special cases of the representation

$$Q = I - YSY^{\mathrm{T}}, \qquad Y \in \mathcal{R}^{m\times k}, \quad S \in \mathcal{R}^{k\times k} \tag{7}$$

for an $m \times m$ orthogonal matrix. With a triangular matrix S, this representation was first introduced as the *compact WY representation* by Schreiber and Van Loan [9], as a way of expressing the product of k Householder matrices in a computationally more advantageous form.

If S is nonsingular and Y is of rank k, then Q acts on the space $\mathcal{R}(Y)^{\perp}$ as the identity and changes every nonzero vector in $\mathcal{R}(Y)$, which we call the active space of Q. From the preceding discussion we see that Jacobi and Householder reflectors have one-dimensional active subspaces, whereas, except for the identity, Givens rotations have two-dimensional active subspaces.

We show in this paper that the representation (7), which we call *the basis-kernel representation*, is a universal representation for *any* orthogonal matrix. This is proved in the next section, and there we also introduce the so-called orthogonality conditions on Y and S, which must be satisfied for the matrix Q of (7) to be orthogonal. We prove further that any orthogonal matrix can be expressed in basis-kernel form with a triangular kernel, and we show how the familiar representation of orthogonal matrices as products of Householder matrices can be readily deduced from this representation. This theory is also used to show that, for an orthogonal matrix Q mapping a matrix A into a matrix B, there is a "minimal" representation of Q in that its associated basis Y has a minimal number of columns. In Section 3 we describe in detail how the basis Y and the kernel S characterize Q. We also derive a canonical form that makes explicit how Q partitions $\mathbf{R}^n$ into a couple of subspaces in which it acts as the identity, a reflector or a rotator. In Section 4 we derive a generalized form, applicable to arbitrary orthogonal matrices, of the Cayley representation [5]. The generalized Cayley form shows that, in a specified active space of dimension k, there are $k(k-1)/2$ degrees of freedom in choosing a nonsymmetric matrix while there is one and only one symmetric matrix. Finally, we comment on our results and outline directions of future research.

2 The Basis-Kernel Representation of Orthogonal Matrices

Theorem 1 *For any $m \times m$ orthogonal matrix Q there exist a full-rank $m \times k$ matrix Y and a nonsingular $k \times k$ matrix S, $k \leq m$, such that*

$$Q := Q(Y, S) = I - YSY^{\mathrm{T}}. \tag{8}$$

Proof. If $I - Q$ is nonsingular, we may choose $Y = I$ and $S = I - Q$. Otherwise, let X and Y be orthonormal bases of $\mathcal{N}(I-Q)$ and $\mathcal{R}(I-Q)$, the null space and range of $I-Q$, respectively. Then,

$$Q = (X, Y) \begin{pmatrix} I & 0 \\ 0 & I - S \end{pmatrix} \begin{pmatrix} X^{\mathrm{T}} \\ Y^{\mathrm{T}} \end{pmatrix},$$

for some orthogonal matrix $I - S$ that has no eigenvalue at 1. Therefore, S is nonsingular and $Q = I - YSY^{\mathrm{T}}$. ∎

As already mentioned in the preceding section, we call $\mathcal{R}(Y)$ of (8) the *active subspace* of Q (which is uniquely defined by Q as to be seen later) and denote it with $\mathcal{A}(Q)$. We define the *degree* of Q as the dimension of $\mathcal{A}(Q)$. We call S the *kernel* of Q, Y the *basis*, and (8) the *basis-kernel representation* of Q. So, for example, a Householder matrix (3) is an orthogonal matrix of degree 1.

Let X_y and X_s be two j-by-k matrices, $j \geq k$, such that $X_y^{\mathrm{T}} X_s = I$. Then, $YSY^{\mathrm{T}} = (YX_y^{\mathrm{T}})(X_s S X_s^{\mathrm{T}})(X_y Y^{\mathrm{T}})$. Hence, a particular orthogonal matrix Q

has many basis-kernel representations of the form of (8), and Y and S need not necessarily be of full rank.

2.1 The Orthogonality Conditions

Like the condition on v and β in (3) for a Householder reflector, there is a condition on Y and S that guarantees the orthogonality of $Q(Y, S)$.

Lemma 2 *1. The orthogonality condition*

$$SY^{\mathrm{T}}YS^{\mathrm{T}} = S + S^{\mathrm{T}} \tag{9}$$

or

$$S^{T}Y^{\mathrm{T}}YS = S + S^{\mathrm{T}} \tag{10}$$

is a sufficient condition for the orthogonality of $Q(Y, S)$.

2. The condition (9) and the condition (10) are equivalent.

3. When S is nonsingular, the orthogonality conditions can be expressed in the unified form

$$Y^{T}Y = S^{-1} + S^{-T}. \tag{11}$$

Proof. Part 1 and 3. If we write $Q = I - YSY^{T}$, then the condition (9) implies $QQ^{\mathrm{T}} = I$ and the condition (10) implies $Q^{\mathrm{T}}Q = I$. The expression of (11) follows immediately from the conditions in Part 1 when S is nonsingular. Part 2. Now assume S is of rank $r < k$. Let $S = U \begin{pmatrix} \Sigma & \\ & 0 \end{pmatrix} V^{\mathrm{T}}$ be a singular value decomposition of S with $\Sigma \in \mathbf{R}^{r\times r}$ nonsingular. Then,

$$U^{\mathrm{T}}SU = \begin{pmatrix} \Sigma & \\ & 0 \end{pmatrix} \begin{pmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{pmatrix} (U_1\ U_2) = \begin{pmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ 0 & 0 \end{pmatrix},$$

where $\tilde{S}_{11} = \Sigma V_1^{\mathrm{T}} U_1$ is a square matrix, and $\tilde{S}_{12} = \Sigma V_1^{\mathrm{T}} U_2$. The orthogonality condition (9) can then be expressed as

$$\begin{pmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ 0 & 0 \end{pmatrix} (YU)^{\mathrm{T}}(YU) \begin{pmatrix} \tilde{S}_{11}^{\mathrm{T}} & 0 \\ \tilde{S}_{12}^{\mathrm{T}} & 0 \end{pmatrix} = \begin{pmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \tilde{S}_{11}^{\mathrm{T}} & 0 \\ \tilde{S}_{12}^{\mathrm{T}} & 0 \end{pmatrix}. \tag{12}$$

The last equation implies that $\tilde{S}_{12} = 0$ and that $\tilde{S}_{11}$ must be nonsingular. Thus, $S = U_1 \tilde{S}_{11} U_1^{\mathrm{T}}$. Multiplying (12) by $\tilde{S}_{11}^{-1}$ and $\tilde{S}_{11}^{-T}$ from the left and right, respectively, we obtain

$$(YU_1)^{\mathrm{T}}(YU_1) = \tilde{S}_{11}^{-1} + \tilde{S}_{11}^{-\mathrm{T}}.$$

Therefore,

$$\tilde{S}_{11}^{\mathrm{T}}(YU_1)^{\mathrm{T}}(YU_1)\tilde{S}_{11} = \tilde{S}_{11} + \tilde{S}_{11}^{\mathrm{T}},$$

and hence the condition (10). In the same fashion, (10) implies $\tilde{S}_{12} = 0$. ■

Given Y, we now show some examples of choices for S such that the orthogonality condition is satisfied.

Example 3 *$Q(Y, S)$ is orthogonal if*

$$S = 2(Y^{\mathrm{T}}Y)^{\dagger},$$

where $B^{\dagger}$ denotes a pseudo-inverse of the matrix B [12].

Such a singular and symmetric kernel was first introduced in [8].

Example 4 *$Q(Y, S)$ is orthogonal if Y has no zero column and*

$$S = [\mathrm{tril}(Y^{\mathrm{T}}Y) + \mathrm{diag}(Y^{\mathrm{T}}Y)/2]^{-1},$$

or

$$S = [\mathrm{triu}(Y^{\mathrm{T}}Y) + \mathrm{diag}(Y^{\mathrm{T}}Y)/2]^{-1},$$

where $\mathrm{tril}(A)$ *(*$\mathrm{triu}(A)$*) is the strictly lower (upper) triangular part of matrix A, and* $\mathrm{diag}(A)$ *is the diagonal of matrix A.*

Note that the triangularity of S and the orthogonality condition (11) together imply that S is unique. One can see that, given Y, the triangular kernel is easy to compute. As a matter of fact, it is the procedure for computing the compact WY representation proposed in [13,7].

2.2 Regularity Assumption

The discussion following Theorem 1 and the examples above have shown that Y and S need not necessarily be of full rank. On the other hand, we know from Theorem 1 that for an orthogonal matrix, there is always a basis-kernel representation with full rank Y and nonsingular S. Such a representation we call a *regular* basis-kernel representation. Under the regularity assumption, the active space of Q is $\mathcal{R}(Y)$. A nonregular basis-kernel transformation can easily be transformed into a regular one, as follows.

Suppose Y is rank deficient. Let $YP = \tilde{Y}R$, with $R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & 0 \end{pmatrix}$, be a rank-revealing QR decomposition of Y (see, for example, [2,3]), that is, R_{11} is nonsingular and $\text{rank}(R_{11}) = \text{rank}(Y)$. Then, $Q(Y,S) = Q(\tilde{Y},\tilde{S})$ with $\tilde{S} = RP^{\mathrm{T}}SPR^{\mathrm{T}}$. Thus, we can assume without loss of generality that Y is of full rank.

Now suppose S is singular. We know from the proof of Lemma 2 that $S = U\bar{S}U^{\mathrm{T}}$ for some U and $\bar{S}$ of full rank. Thus, $Q(Y,S) = Q(\bar{Y},\bar{S})$ with $\bar{Y} = YU$. We therefore assume in the rest of the paper that a basis-kernel representation of an orthogonal matrix is regular unless explicitly stated otherwise.

2.3 Triangular Kernels

For a given Y, the triangular kernel of Example 4 presents another way for computing the compact WY form of a product of Householder reflectors. In fact, *any* orthogonal matrix can be expressed in basis-kernel form with an upper or lower triangular kernel.

Theorem 5 *Any orthogonal matrix Q can be expressed as $Q = I - YSY^T$ with a triangular kernel S.*

Proof. Let $Q = Q(Y,S)$ be an orthogonal matrix of degree k. It is sufficient to prove the claim that there is a (unit) lower matrix L such that $S = L^{\mathrm{T}}RL$ for some upper triangular matrix R, since $Q(YL^{\mathrm{T}}, R)$ will be a basis-kernel representation of Q with triangular kernel. The claim holds for orthogonal matrices of degree $k = 1$. Let $Q(Y,S)$ be an orthogonal matrix of degree $k > 1$. Suppose the claim holds for all matrices of degree less than k. Partition

S^{-1},

$$T = S^{-1} = \begin{pmatrix} \tau & a^{\mathrm{T}} \\ b & \tilde{T}_{-1} \end{pmatrix}.$$

The orthogonality condition (11) implies $2\tau = e_1^{\mathrm{T}}(Y^{\mathrm{T}}Y)e_1 \neq 0$. Thus,

$$L_1 T L_1^{\mathrm{T}} = \begin{pmatrix} \tau & (a-b)^{\mathrm{T}} \\ 0 & T_{-1} \end{pmatrix}, \tag{13}$$

with

$$L_1 = \begin{pmatrix} 1 & 0 \\ -b/\tau & I \end{pmatrix}, \text{ and } T_{-1} = \tilde{T}_{-1} - ba^{\mathrm{T}}/\tau.$$

Substituting (13) into (11) results in

$$L_1(Y^{\mathrm{T}}Y)L_1^{\mathrm{T}} = \begin{pmatrix} \tau & (a-b)^{\mathrm{T}} \\ 0 & T_{-1} \end{pmatrix} + \begin{pmatrix} \tau & 0 \\ (a-b) & T_{-1}^{\mathrm{T}} \end{pmatrix}. \tag{14}$$

Now let

$$Y_{-1} = Y \begin{pmatrix} -b^{\mathrm{T}}/\tau \\ I \end{pmatrix}, \text{ and } S_{-1} = T_{-1}^{-1}.$$

We know from (14) that $I - Y_{-1}S_{-1}Y_{-1}$ is an orthogonal matrix of degree $k-1$. With the induction hypothesis, there is a unit lower triangular matrix L_{-1} and an upper triangular matrix R_{-1} such that $S_{-1} = L_{-1}^{\mathrm{T}} R_{-1} L_{-1}$. With

$$L = \begin{pmatrix} 1 & \\ & L_{-1} \end{pmatrix} L_1, \text{ and } R = \begin{pmatrix} \tau^{-1} & (b-a)^{\mathrm{T}} L_{-1} R_{-1}^{-1} \\ & R_{-1}^{-1} \end{pmatrix},$$

we then have $S = L^{\mathrm{T}} R L$.

Similarly, we can find nonsingular upper triangular matrices R and lower triangular matrices L such that $S = R^{\mathrm{T}} L R$, $S = L^{\mathrm{T}} R^{\mathrm{T}} L$, or $S = R^{\mathrm{T}} L^{\mathrm{T}} R$. The last two decompositions follow from the fact that S^{T} is the kernel of Q^{T}. ∎

Example 4 shows that, for a fixed Y, the upper (lower) triangular kernel is unique. An orthogonal matrix, however, has more than one representation with an upper (lower) triangular kernel. Let $Q(Y, S)$ be a representation with upper triangular kernel S. There is an orthogonal matrix U such that $U^{\mathrm{T}} S U$ is also upper triangular [6, p. 385], and hence $Q(YU, U^{\mathrm{T}} S U)$ is another representation of Q with triangular kernel.

From the compact WY representation we know that the product of k Householder matrices can be expressed in basis-kernel form. The converse holds true as well.

Corollary 6 *Any orthogonal matrix of degree k can be expressed as a product of exactly k nontrivial Householder reflectors.*

Proof. We prove the corollary by induction on the degree k of the orthogonal matrices. The corollary holds for the case of $k = 1$ since an orthogonal matrix of degree 1 is by itself a Householder matrix. Let $k > 1$, and assume that the theorem is true for all orthogonal matrices of degree $\leq k - 1$. Let Q be an orthogonal matrix of degree k and $Q = I - YSY^{\mathrm{T}}$ with an upper triangular kernel S. The orthogonality condition implies

$$S = (\mathrm{triu}(Y^{\mathrm{T}}Y, 1) + \mathrm{diag}(Y^{\mathrm{T}}Y)/2)^{-1}.$$

If we partition Y as $Y = (y, Y_1)$, then

$$S = \begin{pmatrix} s & -sy^{\mathrm{T}}Y_{\text{-}1}S_{\text{-}1} \\ & S_{\text{-}1} \end{pmatrix},$$

and hence

$$Q = I - ysy^{\mathrm{T}} + ysy^{\mathrm{T}}Y_{\text{-}1}S_{\text{-}1}Y_{\text{-}1}^{\mathrm{T}} - YS_{\text{-}1}Y_{\text{-}1}^{\mathrm{T}} = (I - ysy^{\mathrm{T}})(I - Y_{\text{-}1}S_{\text{-}1}Y_{\text{-}1}^{\mathrm{T}}),$$

where $(I - ysy^{\mathrm{T}})$ is a nontrivial Householder matrix and $(I - Y_{\text{-}1}S_{\text{-}1}Y_{\text{-}1}^{\mathrm{T}})$ is an orthogonal matrix of degree $k - 1$ and can be expressed, by the induction hypothesis, as a product of exactly $k - 1$ Householder matrices. ■

Notice how easy it is to determine the representation of Q in terms of Householder matrices from a basis-kernel representation with triangular kernel. The Householder vectors are simply the columns of the basis Y, and the scaling factors are the corresponding diagonal elements of the kernel S. Since the basis-kernel representation with triangular kernel is not unique, the representation of an orthogonal matrix as product of Householder matrices is not unique, either.

Generalizing the proof of Corollary 6, we note the following result for factorization and composition of arbitrary orthogonal matrices in basis-kernel representation with (block) triangular kernel.

Corollary 7

$$Q_1(Y_1,S_1)Q_2(Y_2,S_2) = I - (Y_1,Y_2)\begin{pmatrix} S_1 & -S_1(Y_1^{\mathrm{T}}Y_2)S_2 \\ & S_2 \end{pmatrix}(Y_1,Y_2)^{\mathrm{T}}.$$

Using this formula, one can, for example, quickly assemble random orthogonal matrices in a "binary-tree" like fashion from lower-degree random orthogonal matrices, deriving, in effect, a parallel block version of the Householder-oriented approach by Stewart [11].

2.4 Block Orthogonal Transformations

The following theorem shows that, if there is an orthogonal transformation that transforms an $m \times k$ matrix A into a matrix B, $k < m$, the degree of Q concerned need not be larger than k.

Theorem 8 *Let A and B be two m-by-k matrices, $k < m$. If $B = QA$ for some orthogonal matrix Q, then Q is either of degree no greater than k or can be replaced by an orthogonal factor of its own with degree no greater than k.*

Proof. Let $Q = Q(Y,S)$ be a basis-kernel representation of Q. Suppose the degree of Q is greater than k. Let $Y^TA = U\begin{pmatrix} M \\ 0 \end{pmatrix}$ be a QR-factorization of Y^TA, with $M \in \mathbf{R}^{r\times r}$, where $r \le k$ is the rank of Y^TA. Then $Q(Y,S) = Q(\bar{Y},\bar{S})$, where $\bar{Y} = YU$ and $\bar{S} = U^TSU$. Partitioning $\bar{Y} = [\bar{Y}_1, \bar{Y}_2]$, where $\bar{Y}_1$ is $n \times r$, we then have $\bar{Y}_2^TA = 0$. From the proof of Theorem 5, $\bar{S} = LRL^{\mathrm{T}}$ for some lower triangular matrix L and upper triangular matrix R. Thus, $Q(\bar{Y},\bar{S}) = Q(\tilde{Y},R)$ with $\tilde{Y} = \bar{Y}L$. If we partition $\tilde{Y} = (\tilde{Y}_1,\tilde{Y}_2)$ in the same fashion as $\bar{Y}$ and partition $R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}$ conformingly, Corollary 7 implies

$$Q = Q(\tilde{Y}_1, R_{11})Q(\tilde{Y}_2, R_{22})$$

and $\tilde{Y_2}^{\mathrm{T}}A = 0$. We therefore have

$$B = QA = Q(\tilde{Y}_1, R_{11})A,$$

as claimed. ∎

Not surprising, when a and b are vectors such that $\|a\|_2 = \|b\|_2$, there is always an orthogonal matrix Q of degree 1(i.e., a Householder matrix) such that $b = Qa$.

In matrix computations, the following elimination problem is fundamental. Given an $m \times k$ matrix A, determine an orthogonal matrix Q such that

$$QA = \begin{pmatrix} C \\ 0 \end{pmatrix}, \tag{15}$$

for some k-by-k matrix C. The usual Householder-based approach constructs an orthogonal matrix Q and an upper triangular matrix C in a column-by-column fashion as a product of k Householder vectors. Using the WY representation, one then can deduce a basis-kernel representation with triangular kernel.

Theorem 8 and its proof lead to the following conclusions:

- The elimination problem (15) can be solved with an orthogonal matrix of degree at most k.

- Finding ways to determine orthogonal matrices directly in terms of their basis and kernel (as compared to products of Householder matrices or Givens rotations) seems preferable to arrive at computationally more advantageous procedures.

- The minimal degree of a solution Q to a transformation problem in a k-dimensional subspace could be even lower that k, which would result in a lower-rank, and hence computationally less expensive, Q.

3 Geometric Properties

In the introduction, we reviewed the geometric properties of reflectors "active" in one-dimensional or multidimensional subspaces and of rotators in two-dimensional subspaces. In Section 2, we showed that the basis-kernel representation is a natural approach for representing, composing, and decomposing orthogonal matrices. This section shows that the basis-kernel representation also makes it easy to understand geometric properties of orthogonal matrices.

3.1 The Basis and Active Subspace

The following theorem shows how Y defines the active space and S specifies the transformation in the active subspace.

Theorem 9 *1. $Qx = x \Leftrightarrow x \in \mathcal{R}(Y)^{\perp}$.*

2. For any $u \in \mathcal{R}(Y)$, there exists one and only one vector b such that $u = YS^{\mathrm{T}}b$, and $Qu = -v$, where $v = YSb$.

Proof. Part 1. For any x such that $Qx = x$, we have $YSY^{\mathrm{T}}x = 0$. Since Y has full rank, $YSY^{\mathrm{T}}x = 0$ if and only if $SY^{\mathrm{T}}x = 0$. Thus, $x \in \mathcal{R}(Y)^{\perp}$ if and only S is nonsingular.

Part 2. Since Y is a basis for its own column space and S is nonsingular, any vector u in $\mathcal{R}(Y)$ can be uniquely represented in the form $u = YS^{\mathrm{T}}b$ for some vector b. By the orthogonality condition, we have

$$Qu = (I - YSY^{\mathrm{T}})YS^{\mathrm{T}}b = YS^{\mathrm{T}}b - Y(S + S^{\mathrm{T}})b = -YSb = -v.$$

■

Thus, when $k < m$, the matrix Q has eigenvalues at 1, and the orthogonal complement of $\mathcal{R}(Y)$ is the invariant subspace of Q corresponding to its eigenvalues at 1. Further, on $\mathcal{R}(Y)$, vectors $u = YS^{\mathrm{T}}b$ and $v = YSb$ in $\mathcal{R}(Y)$ are images of each other under the mappings Q and Q^{-1}, respectively.

With respect to the composition of orthogonal matrices, Corollary 7 shows that, if $\mathcal{R}(Y_1) \cap \mathcal{R}(Y_2) = \{0\}$, then $\mathcal{A}(Q_1Q_2) = \mathcal{R}(Q_1) \oplus \mathcal{R}(Q_1)$, or degree$(Q_1Q_2)$ = degree(Q_1) + degree(Q_2). On the other hand, if $Y_2 = Y_1$ and $S_2 = S_1^{\mathrm{T}}$, then the degree of $Q_1Q_2 = I$ is zero. In general, we have the following.

Corollary 10 *Let Q_1 and Q_2 be two orthogonal matrices. Then,*

$$\mathcal{A}(Q_1Q_2) \subseteq \mathcal{A}(Q_1) \oplus \mathcal{A}(Q_2).$$

3.2 The Kernel

While the basis Y determines the space acted upon by Q, the kernel S specifies the action taken in this subspace.

Theorem 11 *1)* $\lambda(Q) = \lambda(-SS^{-\mathrm{T}}) \cup \{1\}$.

2) $\det(Q) = \begin{cases} 1, & \textit{if } k \textit{ is even,} \\ -1, & \textit{otherwise} \end{cases}$

3) $Qx = -x \Rightarrow x \in \mathcal{R}(Y)$ *if and only if* S *is symmetric.*

Proof. Part 1. When S is nonsingular, the orthogonality condition can be expressed as

$$S(Y^{\mathrm{T}}Y) = SS^{-\mathrm{T}} + I.$$

For any vector $y \in \mathcal{R}(Y)$, there exists a unique vector b such that $y = Yb$, and

$$Qy = (I - YSY^{\mathrm{T}})Yb = Yb - Y(SS^{-\mathrm{T}} + I)b = -YSS^{-\mathrm{T}}b. \tag{16}$$

In particular,

$$QY = -Y(SS^{-\mathrm{T}}).$$

By Theorem 9, $\mathcal{R}(Y)$ is the invariant subspace of Q corresponding to all of its eigenvalues not equal to 1. Therefore $\lambda(Q) = \lambda(-SS^{-\mathrm{T}}) \cup \{1\}$.

Part 2. We know from Part 1 that $\det(Q) = \det(-SS^{-\mathrm{T}})$. We then have

$$\det(Q) = (-1)^k \det(S) \det(S^{-1}) = (-1)^k.$$

Part 3. From Part 2 of Theorem 9 and Part 1 of Theorem 11, it remains to show that $Qx = -x$ for any $x \in \mathcal{R}(Y)$ implies that S is symmetric. We see from (16) that

$$\begin{array}{lll} & Qx = -x, & \forall x \in \mathcal{R}(Y), \\ \Leftrightarrow & YSS^{-\mathrm{T}}b = Yb, & \forall\, b \in \mathbf{R}^k, \\ \Leftrightarrow & SS^{-\mathrm{T}} = I & \end{array}$$

and the symmetry of S follows. ■

Note that the determinant of H does not depend on the symmetry of H and that S cannot be skew-symmetric.

Theorem 11 implies that reflectors and symmetric orthogonal matrices are really one and the same.

Corollary 12 *An orthogonal matrix is a reflector iff it is symmetric and not equal to the identity.*

Theorem 11 also illustrates how Q acts upon the subspace $\mathcal{R}(Y)$. The matrix $(-SS^{-T})$ is the representation of Q in $\mathcal{R}(Y)$ with respect to the basis Y, and it has eigenvalues on the unit circle in the complex plane, but not at 1. Let g_j be an eigenvector of $-SS^{-\mathrm{T}}$ corresponding to its eigenvalue $\cos(\theta_j) + i\sin(\theta_j)$. Then,

$$Q(Yg_j) = Y(-SS^{-\mathrm{T}})g_j = (Yg_j)(\cos(\theta_j) + i\sin(\theta_j)).$$

That is, for an arbitrary vector in $\mathcal{R}(Y)$, its components along Yg_j are "rotated" by θ_j, respectively. When Q is a block reflector, the components are rotated uniformly by the same angle π; that is, the sign of vectors in $\mathcal{R}(Y)$ is simply flipped.

If Q should act as other than a reflection on $\mathcal{R}(Y)$, S must be nonsymmetric and $-SS^{-\mathrm{T}}$ must have truly complex eigenvalues, which exist in conjugate pairs. Taking into account Lemma 14, we then have the following corollary,

Corollary 13 *If Q is nonsymmetric, then its kernel S can be expressed with respect to properly chosen Y via*

$$-SS^{-T} = \mathrm{diag}\left(\begin{bmatrix} \cos(\Theta) & \sin(\Theta) \\ -\sin(\Theta) & \cos(\Theta) \end{bmatrix}, B\right), \tag{17}$$

where $B = -I$ or the empty matrix, and $\Theta = \mathrm{diag}(\theta_j)$, $\sin(\theta_j) \neq 0$.

The first diagonal block of (17) can be viewed as a block Givens rotator. Corollary 13 shows that an orthogonal matrix divides its active subspace into two subspaces: it acts as a reflector in one of them and a rotator in the other. An orthogonal matrix of odd degree always has a nontrivial subspace that it acts on as a reflector.

As it turns out, there is a close relationship between SS^{-T} and Y when Y is orthonormal.

Lemma 14 *Let Q be an orthogonal matrix and $Q = I - YSY^{\mathrm{T}}$ be a regular basis-kernel representation of Q. The following statements are equivalent:*

- *Y is orthonormal.*
- *$I - S$ is orthogonal.*

- SS^{-T} *is orthogonal.*

Proof. We have seen from Theorem 1 that if Y is orthonormal, then $I - S$ is orthogonal. Now suppose that $I - S$ is orthogonal. Then $S = I + SS^{-T}$. At the same time, the orthogonality condition (9) implies that

$$S(Y^{\mathrm{T}}Y) = I + SS^{-\mathrm{T}}.$$

Together, they imply that $Y^{\mathrm{T}}Y = I$. ∎

Corollary 13 and Lemma 14 allow us to derive a particularly simple canonical form for S^{-1}.

Theorem 15 *For any orthogonal matrix of degree k there exist an orthonormal basis Y and a kernel S such that*

$$S^{-1} = \frac{1}{2}\begin{pmatrix} I & & \\ & I & D \\ & -D & I \end{pmatrix},$$

where D is either zero or a nonsingular diagonal matrix.

Proof. Let $Q = Q(Y, S)$, and, invoking Corollary 13, assume that Y is orthonormal and (17) holds. From the proof of Lemma 14, we have

$$S^{-1} = (I + SS^{\mathrm{T}})^{-1}.$$

The theorem is true for the special case that $SS^{\mathrm{T}} = I$ with $D = 0$. As another special case consider SS^{T} to be a 2-by-2 Givens rotation $G(\theta)$ with $\sin(\theta) \neq 0$. We then have

$$I+G(\theta) = \begin{pmatrix} 1+\cos(\theta) & \sin(\theta) \\ -\sin(\theta) & 1+\cos(\theta) \end{pmatrix} = 2\cos(\theta/2)\begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix},$$

and since $\sin(\theta) = 2\sin(\theta/2)\cos(\theta/2) \neq 0$,

$$(I + G(\theta))^{-1} = \begin{pmatrix} 1 & -\cot(\theta/2) \\ \cot(\theta/2) & 1 \end{pmatrix}.$$

The claim of the theorem in general easily follows from (17). ∎

4 The Generalized Cayley Representation

For any skew-symmetric matrix B, the matrices

$$(I + B)(I - B)^{-1} \text{ and } (I + B)(B - I)^{-1} \tag{18}$$

are orthogonal. The former does not have eigenvalue at -1, and the latter does not have eigenvalue at 1. Conversely, an orthogonal matrix Q can be represented in one of the above forms with some skew-symmetric matrix B as long as Q does not have eigenvalues at both 1 and -1. Representation (18) is known as the Cayley representation [5] or the Cayley transform of B.

Note that the Cayley representation does not include symmetric orthogonal matrices except I and $-I$, nor does it include the nonsymmetric matrices that have both a nontrivial "inactive" subspace and a nontrivial "active" reflection subspace. We can, however, generalize this representation to cover all orthogonal matrices, by combining the traditional Cayley representation and our basis-kernel representation.

Theorem 16 *Let Y be an orthonormal matrix with k columns. Then Q is an orthogonal matrix with active subspace $\mathcal{R}(Y)$ if and only if*

$$Q = I - Y(I - (B + I)(B - I)^{-1})Y^{\mathrm{T}} \tag{19}$$

for some skew symmetric matrix B. Moreover, Q is symmetric iff $B = 0$.

Proof. It can be checked directly that, for a skew-symmetric matrix B, Q of (19) is orthogonal. On the other hand, if Q is an orthogonal matrix with active subspace $\mathcal{R}(Y)$, then Q can be represented as $Q = I - YSY^{\mathrm{T}}$ for some S that satisfies the equation $I = Y^TY = S^{-1} + S^{-\mathrm{T}}$. Thus, $B = I - 2S^{-1}$ is skew-symmetric and $S = I + SS^{-\mathrm{T}} = I - (B + I)(B - I)^{-1}$. The orthogonal matrix Q is symmetric iff $S = 2I$ and iff $B = 0$. ■

Note that, for the special case that Q has full degree (i.e., no eigenvalue at 1), the generalized Cayley representation (19) becomes the traditional one when one chooses $Y = I$.

Theorem 16 implies that, given a subspace $\mathcal{Y}$ of dimension k, we have $k(k-1)/2$ degrees of freedom in choosing a nonsymmetric orthogonal matrix so that $\mathcal{A}(Q) = \mathcal{Y}$, but there is only one symmetric orthogonal matrix whose active subspace is $\mathcal{Y}$.

5 Conclusions

This paper introduced the basis-kernel representation $Q = I - YSY^T$ of an orthogonal matrix. We showed that any orthogonal matrix can be represented in this form, in particular with a triangular kernel, and showed the relation to the familiar representation of orthogonal matrices as products of Householder matrices.

We also showed how the basis Y determines the subspace that Q acts on in a nontrivial fashion, and how the kernel S determines the action taken on this subspace. This led to a particularly simple representation of $-SS^T$ and S^{-1} which explicitly shows how Q acts on its active subspace as a composition of rotators and reflectors. We also showed that reflectors are exactly the symmetric orthgonal matrices.

Lastly, we generalized the Cayley representation to cover all orthogonal matrices and showed that, given a particular subspace, there is great freedom in choosing nonsymmetric orthogonal matrices acting upon it, but that symmetric orthogonal matrices are uniquely determined by their active subspace.

We also point out that the basis-kernel representation, and the theory we have developed for it, deals directly with Y and S, whereas the usual approaches to orthogonal matrix computations deal principally with elementary operations such as Givens reflectors, Jacobi rotators, or Householder reflectors. Thus, we believe that this representation has profound implications for numerical computations, in that it opens the door to different approaches for deriving orthogonal matrices with desired properties. For example, the proof of Theorem 8 hinted at the possibility for finding lower-rank orthogonal matrices for block elimination problems than the orthogonal matrices provided by the usual approaches.

Acknowledgment

We thank Beresford Parlett for some stimulating discussions.

References

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. DuCroz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK User's Guide*. SIAM, Philadelphia, 1992.

[2] Christian H. Bischof and Per Christian Hansen. Structure-preserving and rank-revealing QR factorizations. *SIAM Journal on Scientific and Statistical Computing*, 12(6):1332–1350, November 1991.

[3] Tony F. Chan. An improved algorithm for computing the singular value decomposition. *ACM Transactions on Mathematical Software*, 8:72–83, 1982.

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

[5] F. R. Gantmacher. *Applications of the Theory of Matrices*. Interscience Publications, Inc., 1959.

[6] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins Press, Baltimore, 2nd edition, 1989.

[7] Chiara Puglisi. Modification of the Householder method based on the compact WY representation. *SIAM Journal on Scientific and Statistical Computing*, 3(3):723–726, 1992.

[8] Robert Schreiber and Beresford Parlett. Block reflectors: Theory and computation. *SIAM Journal on Numerical Analysis*, 25(1):189–205, 1988.

[9] Robert Schreiber and Charles Van Loan. A storage efficient WY representation for products of Householder transformations. *SIAM Journal on Scientific and Statistical Computing*, 10(1):53–57, 1989.

[10] G. W. Stewart. *Introduction to Matrix Computation*. Academic Press, Inc., New York, 1973.

[11] G. W. Stewart. The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM Journal on Numerical Analysis*, 17:403–409, 1980.

[12] G. W. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, Inc., New York, 1991.

[13] Homer F. Walker. Implementation of the GMRES method using Householder transformations. *SIAM Journal on Scientific and Statistical Computing*, 9(1):152–163, 1988.