On Orthogonal Block Elimination*

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

Argonne Preprint MCS-P450-0794

Abstract. We consider the block elimination problem $Q\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} -C \\ 0 \end{pmatrix}$, where, given a matrix $A \in \mathbf{R}^{m\times k}$, $A_{11} \in \mathbf{R}^{k\times k}$, we try to find a matrix C with $C^TC = A^TA$ and an orthogonal matrix Q that eliminates A_2. Sun and Bischof recently showed that any orthogonal matrix can be represented in the so-called basis-kernel representation $Q = Q(Y,S) = I - YST^T$. Applying this framework to the block elimination problem, we show that there is considerable freedom in solving the block elimination problem and that, depending on A and C, we can find $Y \in \mathcal{R}^{m\times r}$, $S \in \mathcal{R}^{r\times r}$, where r is between rank(A_2) and k, to solve the block elimination problem. We then introduce the canonical basis $Y = \begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix}$ and the canonical kernel $S = (A_1 + C)^\dagger C^{-T}$, which can be determined easily once C has been computed, and relate this view to previously suggested approaches for computing block orthogonal matrices. We also show that the condition of S has a fundamental impact on the numerical stability with which Q can be computed and prove that the well-known compact WY representation approach, employed, for example, in LAPACK, leads to a well-conditioned kernel. Lastly, we discuss the computational promise of the canonical basis and kernel, in particular in the sparse setting, and suggest pre- and postconditioning strategies to ensure that S can be computed reliably and is wellconditioned.

Key words. Orthogonal Matrices, Block Elimination, Basis-Kernel Representation, Canonical Basis and Kernel, Householder Matrices, Sparse Matrices, Orthogonal Factorization.

*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 #21, available via anonymous ftp to `ftp.super.org` in the directory pub/prism.

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, [9, 15]).

For dense matrices, orthogonal matrices are most commonly composed via Householder reflectors, which are orthogonal matrices of the form

$$H = H(v) = I - \beta v v^{\mathrm{T}}, \qquad \beta v^{\mathrm{T}} v \beta = 2\beta. \tag{1}$$

The condition on the Householder vector v and the scaling factor β in (1) 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 of v. Computationally, the application of H (or H^T) to a matrix B involves a matrix-vector product and a rank-one update.

On modern machines with memory hierarchies and, in particular, parallel machines with distributed memories, matrix-matrix operations, especially matrix-matrix multiply [7, 8], significantly outperform matrix-vector operations. As a result, there has been considerable interest in so-called block Householder transformations, which express a product

$$Q = H_1 \cdots H_k$$

of several Householder matrices (acting in $\mathcal{R}^m$, say) in a form that allows for the use of matrix-matrix kernels in the application of Q. An example is the *WY representation* [4]

$$Q = I + WY^{\mathrm{T}}, \tag{2}$$

where W and Y are $m \times k$ matrices and Y is composed of the Householder vectors of H_i. Mathematically equivalent is the *compact WY representation* [14]

$$Q = I - YSY^{\mathrm{T}}, \tag{3}$$

where S is a k-by-k triangular matrix. We see that the compact WY representation requires only $O(k^2)$ workspace for the matrix S, compared with the WY representation's $O(mk)$ workspace for W. This reduction in memory requirement may be significant, since typically $m \gg k$. Block orthogonal transformations based on the compact WY form have, for example, been incorporated into the LAPACK library [1].

Such blocking approaches are particular solutions to the *orthogonal block elimination problem*:

$$\boxed{\begin{array}{l} \text{Given a matrix } A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix},\ A \in \mathcal{R}^{m\times k},\ A_1 \in \mathcal{R}^{k\times k},\ m > k. \\ \text{Find an orthogonal matrix } Q \text{ such that } QA = \begin{pmatrix} -C \\ 0 \end{pmatrix} \text{ for some} \\ \text{matrix } C \in \mathcal{R}^{k\times k}. \end{array}} \tag{4}$$

The canonical elimination problem is formulated in terms of $-C$ for notational convenience, as will become evident later on. We call the matrix C the *image of* A under Q. The orthogonality of Q implies that C must satisfy what Schreiber and Parlett [13] called the *isometry property*,

$$C^T C = A^T A. \tag{5}$$

The Cholesky factor of $A^T A$ or the square root of $A^T A$, for example, satisfies this condition. In the former case, C is triangular; in the latter case, it is symmetric.

The usual block Householder approach (see, for example, [9, pp. 211–213]), employed, for example, in LAPACK, to solve the block elimination problem (4) essentially consists of two parts. First, compute an unblocked QR factorization of A to generate k Householder transformations. Second, accumulate a compact WY representation [9, pp. 211–213] for block updates. This approach results in a triangular image C.

Recently, Sun and Bischof [16] showed that, far from being just a convenient way of formulating products of Householder matrices, *any* orthogonal matrix can be expressed in the form (3), where S need not necessarily be triangular. They called this form the *basis-kernel representation of* Q, motivated by the fact that the *basis* Y displays the *active subspace* $\mathcal{R}(Y)$, that is, the subspace where the transformation determined by Q acts in a nontrivial fashion, whereas the *kernel* S determines how the the active subspace is transformed. Among other results, the paper [16] showed the following:

- Given an arbitrary basis-kernel representation of Q, one can constructively derive a *regular* basis-kernel representation, namely, one in which both Y and S have full rank. Under the assumption that Y has full rank, the dimension of $\mathcal{R}(Y)$ is called the *degree* of Q.

- An orthogonal matrix of degree k is sufficient to solve the elimination problem (4).

- Given an arbitrary basis-kernel representation of Q, one can constructively derive a representation with a triangular kernel. However, even regular basis-kernel representations with triangular kernels are not unique.

- The *orthogonality conditions*

  $$SY^{\mathrm{T}}YS^{\mathrm{T}} = S + S^{\mathrm{T}} \tag{6}$$

  or

  $$S^{T}Y^{\mathrm{T}}YS = S + S^{\mathrm{T}} \tag{7}$$

  are sufficient conditions for the orthogonality of $Q(Y, S)$.

- Any orthogonal matrix Q of degree k can be expressed as the product of exactly k nontrivial Householder reflectors, and if $Q = I - YSY^{T}$, with triangular S, then $Q = H_1 \cdots H_k$, where $H_i = I - s_{ii}Y(:, i)Y(:, i)^{T}$.

- If $Q_i = I - Y_iS_iY_i^{T}$, then

  $$Q_1(Y_1, S_1)Q_2(Y_2, S_2) = I - (Y_1, Y_2)\begin{pmatrix} S_1 & -S_1(Y_1^{\mathrm{T}}Y_2)S_2 \\ & S_2 \end{pmatrix}(Y_1, Y_2)^{\mathrm{T}}. \tag{8}$$

Hence, without loss of generality, we restrict orthogonal matrices to the form

$$Q = Q(Y, S) = I = YSY^{T}$$

and require that any Q solving the orthogonal block elimination problem (4) be of degree not more than k. Note that in the particular case $k = 1$ the orthogonal matrix to be determined is a Householder matrix.

In this paper, we investigate the orthogonal block elimination problem employing the framework of the basis-kernel representation in our study. In the next section, we show that, for given C, there is a unique orthogonal matrix with minimum degree, equal to rank $\begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix}$, that solves the block elimination problem. We then introduce in Section 3 the so-called canonical basis $Y = \begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix}$ and the canonical kernel $S = (A_1 + C)^\dagger C^{-T}$ and show that they provide a representation of the unique minimum-degree orthogonal transformation that solves the block elimination problem for a given image of A under Q. In Section 4 we consider a transformed block elimination problem where A has orthonormal columns and use it to prove that the minimum degree achievable for Q over all choices of C is equal to the rank of A_2. We also relate our framework to Schreiber and Parlett's block reflectors and the compact WY accumulation procedure. In Section 5 we show that the conditioning of the kernel S and the scaling of Y's columns has a profound effect on the numerical accuracy with which an orthogonal matrix can be computed. We also show that, using the conventional choice for the computation of Householder vectors, the compact WY accumulation procedure results in very well-conditioned kernels. In Section 6 we discuss the computational advantages of the canonical basis and kernel with respect to the preservation of data locality and exploitation of sparsity, and suggest conditioning approaches to ensure that basis and kernel can be computed accurately. Lastly, we summarize our results.

2 Minimum-Degree Transformations

Suppose, for $m \times k$ matrices A and B, that Q is an orthogonal matrix such that $QA = B$. Let $Q(Y, S) = I - YSY^T$ be a regular basis-kernel representation of Q. Then $A - B = YSY^T$, and the basis Y must satisfy the *inclusion property*

$$A - B \subset \mathcal{R}(Y). \tag{9}$$

The rank of Y, and hence the degree of Q, can therefore be no less than $\operatorname{rank}(A - B)$. We can in fact prove the following theorem.

Theorem 1 *Let A and B be $m \times k$ such that $QA = B$ for some orthogonal matrix Q. Let $r = \operatorname{rank}(A - B)$. Then there exists a unique orthogonal matrix Q_{min} of degree r such that $Q = Q_{min}Q_{null}$ and $Q_{min}A = B$.*

Proof. Let $Q(Y,S) = I - YSY^T$ be a regular basis-kernel representation of Q. Then, $A - B = YSY^{\mathrm{T}}A$ and

$$\mathrm{rank}(Y^{\mathrm{T}}A) = \mathrm{rank}(A-B) =: r.$$

Now consider the case that the degree of Q is greater than r; otherwise $Q_{min} = Q$ and $Q_{null} = I$. Let $Y^T A = U \begin{pmatrix} M \\ 0 \end{pmatrix}$ be a QR-factorization of $Y^T A$, with $M \in \mathcal{R}^{r\times k}$. Then $Q(Y,S) = Q(\bar{Y},\bar{S})$, where $\bar{Y} = YU$ and $\bar{S} = U^T SU$. Partitioning $\bar{Y} = [\bar{Y}_1, \bar{Y}_2]$, where $\bar{Y}_1$ is $m \times r$, we have $\bar{Y}_2^T A = 0$. The proof of Theorem 5 in [16] showed that there exist a lower triangular matrix L and an upper triangular matrix R such that $\bar{S} = LRL^{\mathrm{T}}$. 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, Equation (8) implies

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

and $\tilde{Y_2}^{\mathrm{T}} A = 0$. With

$$Q_{min} = Q(\tilde{Y}_1, R_{11}), \text{ and } Q_{null} = Q(\tilde{Y}_2, R_{22}),$$

we then have

$$Q = Q_{min}Q_{null} \text{ and } B = Q_{min}A\,,$$

as claimed.

Now assume there are two orthogonal matrices Q_1 and Q_2 of degree r that satisfy $Q_i A = B$, and let $Q_i = I - Y_i S_i^T Y_i^T$ be a regular basis-kernel representation with $Y_i \in \mathcal{R}^{m\times r}$, $S_i \in \mathcal{R}^{r\times r}$, and $\mathrm{rank}(Y) = \mathrm{rank}(S) = r$. Let X be an $k \times r$ matrix of rank r such that $Y_1^{\mathrm{T}} AX = I_{r\times r}$. Then, $Y_1 S_1 = Y_2 S_2 (Y_2^{\mathrm{T}} AX)$ and, since $S_2(Y_2^{\mathrm{T}} AX) \in \mathcal{R}^{r\times r}$ has full rank, $Y_2 = Y_1 F$ for a nonsingular matrix F. Since Y_1 is of full rank, we have $S_1 = FS_2(F^{\mathrm{T}} Y_1^{\mathrm{T}} AX) = BS_2 B^{\mathrm{T}}$, which implies $Q_2 = Q_1$. ∎

From the proof of Theorem 1 and the fact that A and B have symmetrical positions ($QA = B$ implies $Q^T B = A$), we can deduce the following facts.

Corollary 2 *Let $QA = B$, where $Q = I - YSY^{\mathrm{T}}$ is a regular basis-kernel representation of Q.*

1. *The following statements are equivalent:*
   - $Q(Y,S)$ *is of minimum degree.*
   - $R(Y) = R(A-B)$.
   - $Y^T A$ *is of full rank in rows.*
   - $Y^T B$ *is of full rank in rows.*
2. *If* $Q(Y,S)$ *is not of minimum degree, then it can be factored as*

$$Q(Y,S) = Q_1(Y_1,S_1)\,Q_2(Y_2,S_2)\,,$$

where

$$\mathcal{R}(Y_1) = \mathcal{R}(A-B), \;\; \textit{and}\; Q_1(Y_1,S_1)A = B$$

and

$$\mathcal{R}(Y_2) \subset \mathcal{R}(A)^{\perp}, \;\; \textit{or}\; \mathcal{R}(Y_2) \subset \mathcal{R}(B)^{\perp}.$$

In our study of the block elimination problem (4), we consider the particular case $B = \begin{pmatrix} -C \\ 0 \end{pmatrix}$, where C is a $k \times k$ matrix that satisfies the isometry condition (5). Once C is determined, we know that the minimum degree for a solution to the block elimination problem is the rank of $\begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix}$. Note that the minimum degree discussed in Theorem 1 depends on the chosen image C. In particular, if we can choose C such that $\begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix}$ does have a rank lower than k, we may be able to arrive at a more economical representation of Q, in the sense that the computational cost of applying Q in the basis-kernel form is directly proportional to the number of columns of Y.

3 The Canonical Basis and Kernel

Let $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ be an $m \times k$ $(m > k)$ matrix where A_1 is $k \times k$, and let C by a $k \times k$ matrix that satisfies the isometry condition with A. By Theorem 1, there exists an orthogonal matrix Q of degree equal to the rank of $\begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix}$ that solves the block elimination problem (4). We now show that a particular basis Y and kernel S representing Q (i.e., $Q = Q(Y, S)$) can be derived directly from A and C, without any need for a column-by-column approach as is, for example, employed in the WY approaches. In this section, we assume that A is of full rank. The rank-deficient case is addressed in Section 6.

We have already established that Y must satisfy the inclusion property (9)

$$\mathcal{R}(\begin{bmatrix} A_1 + C \\ A_2 \end{bmatrix}) \subseteq \mathcal{R}(Y).$$

Also note that $A = Q(Y, S)^T \begin{pmatrix} -C \\ 0 \end{pmatrix}$ implies that Y should satisfy the *elimination condition*

$$Y = Y S^{\mathrm{T}} (A_1 + C)^{\mathrm{T}} C, \tag{10}$$

which suggests (and if Y is of full rank, actually implies)

$$S^{\mathrm{T}} (A_1 + C)^{\mathrm{T}} C = I\,.$$

In fact, we can prove the following theorem.

Theorem 3 *Let $A \in \mathcal{R}^{m \times k}$, $m > k$, be of full rank, and let $C \in \mathcal{R}^{k \times k}$ satisfy the isometry condition with A. Then $Q(Y, S)$, defined by*

$$Y = A + \begin{pmatrix} C \\ 0 \end{pmatrix} = \begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix} \tag{11}$$

and

$$S = (A_1 + C)^{\dagger} C^{-T}, \tag{12}$$

solves the block elimination problem $Q(Y, S)A = \begin{pmatrix} -C \\ 0 \end{pmatrix}$. Furthermore, Y and S provide a basis-kernel representation for the unique orthogonal matrix of minimum degree that solves the block elimination problem.

Proof. It can be verified directly that Y and S as defined above satisfy the orthogonality condition (6), and therefore $Q(Y, S) = I - Y S Y^T$ is orthogonal.

If (A_1+C) is nonsingular, it is easy to check that Y and S solve the block elimination problem.

Now consider the case where $A_1 + C$ has rank $r < k$. From the first k rows of the elimination condition (10), we have

$$C^{\mathrm{T}}(A_1 + C) = C^{\mathrm{T}}(A_1 + C)S^{\mathrm{T}}(A_1 + C)^{\mathrm{T}}C.$$

Using a singular value decomposition, we have $C^{-1}(A_1 + C) = V_L \Sigma V_R^{\mathrm{T}}$ for a nonsingular matrix $\Sigma \in \mathcal{R}^{r\times r}$ and orthonormal matrices V_L and V_R. Thus,

$$V_L \Sigma V_R^{\mathrm{T}} = V_L \Sigma V_R^{\mathrm{T}} S^{\mathrm{T}} V_R \Sigma V_L^{\mathrm{T}}.$$

We therefore know that there exists a nonsingular $r \times r$ matrix B such that $C^{\mathrm{T}}(A_1 + C) = V_L B V_L^{\mathrm{T}}$. Since $(A_1 + C)^{\dagger} C^{-T} = V_L B^{-1} V_L^{\mathrm{T}}$, it is easy to check that Y as in (11) and S as in (12) together satisfy the elimination condition (10).

Independent of the regularity of $A_1 + C$, we have by construction

$$\mathcal{R}(Y) = \mathcal{R}\left(A + \begin{pmatrix} C \\ 0 \end{pmatrix}\right),$$

and so, according to Corollary 2, $Q(Y,S)$ has minimum degree. ■

We call (11) and (12) the "canonical base" and the "canonical kernel", respectively. Note that, for $k = 1$, these formulae are exactly those defining for a Householder transformation. It is also worth pointing out that, for a given C, Y and S need not be of full rank, even though $Q(Y,S)$ employing the canonical base and kernel is the unique orthogonal transformation of minimum degree solving the block elimination problem.

Corollary 4 *Let Y and S be defined as in Theorem 3. Then, Y and S are of full rank if and only if $(A_1 + C)$ is nonsingular.*

Proof. The elimination condition (10) implies that if Y is of full rank, then the matrix product $S^{\mathrm{T}}(A_1 + C)^{\mathrm{T}}C$ must be nonsingular. This, in turn, implies the nonsingularity of $A_1 + C$ and S. On the other hand, since $A_1 + C$ is the first k-by-k submatrix of Y, the nonsingularity of $A_1 + C$ implies the independence of Y's columns. ■

Note that, for a particular chosen image C, A_2 plays no role in determining the minimum degree required to solve the block elimination problem.

4 Orthogonal Factors of Images

In the preceding sections we assumed that the image C of A under Q had been fixed a priori, and we showed how one could easily derive the canonical basis and kernel that solved the block elimination problem for that particular choice of C. In this section, we consider the impact of the choice of C.

Assume again that A is of full rank. Let Q and $\overline{Q}$ be two orthogonal transformations that both eliminate A_2, namely,

$$QA = \begin{pmatrix} -C \\ 0 \end{pmatrix}, \quad \text{and} \quad \overline{Q}A = \begin{pmatrix} -\overline{C} \\ 0 \end{pmatrix}. \tag{13}$$

Since both C and $\overline{C}$ satisfy the isometry condition with A, we have $C^TC = \overline{C}^T\overline{C}$ and hence $\overline{C} = UC$ for some $k \times k$ orthogonal matrix U.

Let $G = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}$ be the first k columns of Q^{T}. Then $A = -GC$ and $\mathcal{R}(G) = \mathcal{R}(A)$. Equation (13) implies

$$QA = \begin{pmatrix} -C \\ 0 \end{pmatrix}, \qquad QG = \begin{pmatrix} -I \\ 0 \end{pmatrix}, \tag{14}$$

and

$$\overline{Q}A = \begin{pmatrix} -UC \\ 0 \end{pmatrix}, \qquad \overline{Q}G = \begin{pmatrix} -U \\ 0 \end{pmatrix}. \tag{15}$$

That is, if we fix an orthonormal basis G of $\mathcal{R}(A)$ and hence the representation of A, namely $-C$, with respect to the basis, the block elimination problem (4) is mathematically equivalent to the following problem, which we call the transformed block elimination problem:

> Given a matrix $G = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} \in \mathcal{R}^{m\times k}$, $G_1 \in \mathcal{R}^{k\times k}$, $G_1^TG_1 + G_2^TG_2 = I$, and an orthogonal matrix U, find an orthogonal matrix $Q = I - YSY^T$ of degree no more than k such that $Q\begin{pmatrix} G_1 \\ G_2 \end{pmatrix} = \begin{pmatrix} -U \\ 0 \end{pmatrix}$. (16)

For a solution $Q(Y,S)$ of this transformed problem, we have

$$Q(Y,S)A = Q\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} -UC \\ 0 \end{pmatrix}. \tag{17}$$

Note the strong resemblance to the Householder point elimination problem,

$$H(v, \tau)a = \begin{pmatrix} -\gamma\|a\|_2 \\ 0 \end{pmatrix},$$

where $\gamma = \pm 1$. Partitoning $a = \begin{pmatrix} \alpha \\ a_2 \end{pmatrix}$, $\alpha \in \mathcal{R}$, we choose γ in LAPACK as

$$\gamma = \begin{cases} \mathrm{sgn}(\alpha_1), & \text{if } a_2 \neq 0, \\ -\mathrm{sgn}(\alpha_1), & \text{otherwise.} \end{cases} \tag{18}$$

4.1 Degree Choices

Applying the results in Section 3, we know that, given U,

$$Y = \begin{pmatrix} G_1 + U \\ G_2 \end{pmatrix}, \qquad S = (G_1 + U)^{\dagger} U$$

provide a representation $Q(Y, S)$ for the unique orthogonal matrix Q of minimum degree that solves the problem (16) for that particular choice of U. We also know from Corollary 2 that $r := \mathrm{rank}(Y^T G) = \mathrm{rank}(G_1^T U + I)$ is the degree of Q, with the inclusion property (9) implying $r \leq k$.

The following lemma shows that there is always a $k \times k$ matrix U such that $G_1 + U$ is nonsingular. That is, for this particular choice of U the minimum degree for the related elimination problem is k.

Lemma 5 *Let $G_1 = UR_1$ be a QR decomposition of G_1 scaled such that R_1 has nonnegative diagonal elements. Then the matrix $G_1 + U$ is nonsingular, and the canonical kernel $S = (U^{\mathrm{T}} G_1 + I)^{-1}$ is upper triangular.*

Proof: Since R has nonnegative diagonal elements, $G_1 + U = U(R + I)C$ is nonsingular. The triangularity of S follows easily from the choice of U. ■

This fact was intuitively expected, as the triangularization of a matrix with k columns usually requires k Householder transformations. We now show how the rank of G_2 relates to the lowest degree we can achieve through proper choice of U.

Lemma 6 *For any orthogonal factor U,*

$$\mathrm{rank}(G_1 + U) = \mathrm{rank}(\begin{pmatrix} G_1 + U \\ G_2 \end{pmatrix}).$$

Proof. The claim holds if $G_1 + U$ is nonsingular. We consider the case that $G_1 + U$ is singular. From the proof of Lemma 2 in [16], we have

$$U^{\mathrm{T}}(G_1 + U) = V_s \begin{pmatrix} \overline{S} & \\ & 0 \end{pmatrix} V_s^{\mathrm{T}}$$

for some orthogonal matrix V_s and a nonsingular matrix $\overline{S}$ of order, say, r. Thus,

$$G_1 V_s = U V_s \begin{pmatrix} \overline{S} - I & \\ & -I \end{pmatrix}.$$

The CS decomposition of G (see, for example, [9]) then implies that the last $k - r$ columns of $G_2 V_s$ are zeros exactly when the last $k - r$ columns of $(G_1 + U)V_s$ are zero and its first r columns are of full rank. That is,

$$\mathrm{rank}(\begin{pmatrix} G_1 + U \\ G_2 \end{pmatrix} V_s) = \mathrm{rank}((G_1 + U)V_s).$$

$\blacksquare$

Theorem 7 *There exists an orthogonal matrix U such that $G_1 + U$ is singular if and only if G_2 is rank deficient. Moreover,*

$$\mathrm{rank}(G_2) = \min\{\mathrm{rank}(G_1 + U) \,|\, U \text{ is orthogonal}\}.$$

Proof. Suppose $G_1 + U$ is singular for some orthogonal matrix U. The proof of Lemma 6 shows that G_2 is rank deficient. Now assume $r = \mathrm{rank}(G_2) < k$. We have from Lemma 6 that for any orthogonal factor U,

$$\mathrm{rank}(G_1 + U) = \mathrm{rank}(\begin{pmatrix} G_1 + U \\ G_2 \end{pmatrix}) \geq \mathrm{rank}(G_2).$$

By the CS decomposition theory, there are three orthogonal matrices V_1, V_2, and V_r such that

$$G_1 = V_1 \begin{pmatrix} I & \\ & \Sigma_1 \end{pmatrix} V_r^{\mathrm{T}}, \text{ and } G_2 = V_2 \begin{pmatrix} 0 & \\ & \Sigma_2 \end{pmatrix} V_r^{\mathrm{T}},$$

where Σ_i are nonsingular diagonal matrices and $\Sigma_1^2+\Sigma_2^2 = I$. Let $U = -V_1 V_r$. Then $G_1 + U$ is singular and

$$\mathrm{rank}(G_1 + U) = \mathrm{rank}((G_1^{\mathrm{T}} + U^{\mathrm{T}}, G_2^{\mathrm{T}})) = \mathrm{rank}(G_2).$$

∎

Theorem 7 reveals that the minimum degree for orthogonal transformations that solve the block elimination problem depends on the rank of G_2, or A_2, the block to eliminate. Note the difference between the lowest degree of orthogonal transformations with a *given* image U (Theorem 1) and the minimum degree of orthogonal transformations for eliminating G_2 among all possible choices in U (Theorem 7). That is, depending on the choice of the image of A, the degree of the orthogonal matrix for the block elimination problem for a full rank matrix A with k columns may be anywhere between $\mathrm{rank}(A_2)$ and k.

4.2 Block Reflectors

Schreiber and Parlett [13] developed a theory on block reflectors, which are symmetric orthogonal matrices

$$Q = I - YTY^{\mathrm{T}}, \qquad Q^{\mathrm{T}}Q = I, \qquad T = T^{\mathrm{T}}, \quad Y \in \mathcal{R}^{m\times k}. \tag{19}$$

A particular example is the situation where Y has orthonormal columns and $T = 2I$. From the discussion in preceding sections, we know that *all* block orthogonal transformations — hence, in particular, symmetric ones — solving the block elimination problem (17) can be characterized as special choices of the orthogonal factor U.

Example 8 *(***Block Reflectors***). Suppose $G_1 = V_l \Sigma V_r^{\mathrm{T}}$ is a SVD of G_1. Let $U = V_l D V_r^{\mathrm{T}}$ for some (real) diagonal matrix D such that $|D| = I$. Then*

1. *$G_1 = UM$ is a polar decomposition of G_1 [9, p. 148], where $M = V_r \Sigma D V_r^{\mathrm{T}}$ is symmetric.*
2. *$S = (G_1 + U)^{\dagger} U = (I + M)^{\dagger} V = V_r (I + \Sigma D) V_r^T$ is symmetric.*
3. *If, in addition, D is chosen such that $I + \Sigma D$ is positive definite (which is always possible), then S is positive definite, and $Q = Q(Y, S)$ can be represented in the special form of*

$$Q = I - \tilde{Y}\tilde{Y}^{\mathrm{T}}, \quad \textit{with} \quad \tilde{Y} = YS^{1/2}.$$

Example 8 shows that many symmetric orthogonal transformations solve the block elimination problem. They result in different images of A, and the condition of their kernels differs. The diagonal matrix D in Example 8 can be chosen to minimize, say, the two-norm condition number of S, as

$$\kappa_2(S) = \max\left\{\frac{|\sigma_i + d_i|}{|\sigma_j + d_j|} \text{ such that } \sigma_j + d_j \neq 0, 1 \leq i \leq k, 1 \leq j \leq k\right\}.$$

For the case $k = 1$, $U = d = \pm 1$ and $S = (1 \pm g_1 d)^{-1}$. The LAPACK selection $d = \text{sign}(g_1)$ (see (18)) for a nontrivial Householder matrix results in the smaller scaling factor S among the two choices. Parlett [12] showed, however, that the alternate can be computed in an equally stable fashion.

In their computational procedures for block reflectors, Schreiber and Parlett [13] use Higham's algorithm [10] to compute the polar decomposition of G_1, and hence implicitly chose $D = I$. In this case, $S = V_r(I + \Sigma)^{-1}V_r^T$ is positive definite, and S is always extremely wellconditioned, as $\kappa_2(S) \leq 2$. The case that the actual block size could be smaller than the number of columns of A was first mentioned in [13], although the link to the rank of A_2 (or G_2) was not recognized.

4.3 The Compact WY Representation

The WY approach for generating orthogonal transformations for the block elimination problem does not require an orthonormal basis of $\mathcal{R}(A)$ or of $\mathcal{R}(Y)$. Denote by $\text{WY}(y_1, \cdots, y_k) = I - YSY^T$ the compact WY representation for the product of Householder transformations as derived from the conventional QR factorization, where $Y = (y_1, \cdots, y_k)$ is composed of the Householder vectors y_i, and a null vector denotes a degenerate Householder matrix (i.e., the identity) resulting from a "skipped" orthogonal transformation.

Applying the conventional WY approach to the transformed block elimination problem (16), we get a diagonal image U as a result of the orthogonality among G's columns. The diagonal elements of U are determined one by one by the rule (18). On the other hand, Corollary 6 in [16] showed that, given an arbitrary $k \times k$ diagonal matrix D, $|D| = I$, one can determine a sequence of Householder matrices with corresponding Householder vectors

$y_1, \cdots, y_k$ so that

$$\mathrm{WY}(y_1, \cdots, y_k)G = \begin{pmatrix} -D \\ 0 \end{pmatrix}. \tag{20}$$

As it turns out, orthogonal transformations for the block elimination problem generated with the WY approach have minimum degree.

Theorem 9 *Suppose* $WY(y_1, \cdots, y_k)G = \begin{pmatrix} -D \\ 0 \end{pmatrix}$ *for a real diagonal matrix* D, $|D| = I$. *Then* $WY(y_1, \cdots, y_k)$ *is the minimum-degree transformation that solves the transformed block elimination problem (16) with* $U = D$.

Proof. If the Householder matrices determining $\mathrm{WY}(y_1, \cdots, y_k)$ are all equal to the identity, then $WY(y_1, \cdots, y_k) = I$ is already of the lowest degree. Otherwise, it was shown in [16] that $WY(y_1, \cdots, y_k) = Q(\hat{Y}, \hat{S})$ for a regular basis $\hat{Y}$ and a nonsingular kernel $\hat{S}$, where $\hat{Y}$ consists of the nonzero Householder vectors. Since the first nonzero elements of these Householder vectors occur in different rows, $(D, 0)\hat{Y}$ is of full column rank. The claim in the theorem then follows from Corollary 2. ∎

Assume we have fixed the factor C in A's image to be upper triangular. By the uniqueness of minimum degree transformations, we can then deduce the following fact.

Corollary 10 *The class of orthogonal matrices determined by the WY approach for the block elimination problem (4) with triangular images* C *is the same class of minimum-degree orthogonal matrices that solves the transformed block elimination problem (16) with diagonal images* U.

Recall that, given a particular C, the canonical basis and kernel also provide the minimum degree orthogonal transformation for the problem (4). The WY approach is another way to compute such a block transformation, choosing Y column by column.

Again, we point out that the minimum-degree orthogonal transformation associated with a particular image is not necessarily the minimum-degree transformation possible overall. As an example, consider the matrix

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \beta & \beta \end{pmatrix} \in \mathcal{R}^{3\times 2},$$

where $\alpha_{ij} \in \mathcal{R}$ and $\beta \neq 0$. Two nontrivial Householder matrices, and hence an orthogonal transformation of degree 2, will be required unless $(\alpha_{11} \pm \|Ae_1\|_2)(\alpha_{12} - \alpha_{11}) = \alpha_{21}(\alpha_{21} - \alpha_{22})$. By Theorem 7, on the other hand, the lowest degree for the block elimination problem is 1, the rank of A's submatrix (β, β).

5 The Condition of the Kernel

In the preceding two sections, we have introduced the canonical basis and kernel for a representation of minimum-degree orthogonal transformations to the block elimination problem. We have also shown the relation of the compact WY accumulation procedure with diagonal images. However, mathematically equivalent representations may have quite different numerical properties, and we now discuss to what extent the kernel influences the numerical properties of particular representations of orthogonal matrices.

Theorem 11 *Let Y be an orthogonal basis of full rank and S be its associated kernel. Suppose $\tilde{S} = S + \Delta S$ is a kernel computed from Y. Let $\tilde{Q} = I - Y\tilde{S}Y^{\mathrm{T}}$. Then*

$$\|\tilde{Q}^{\mathrm{T}}\tilde{Q} - I\|_F \leq 4\|S^{-1}\|_F\|\Delta S\|_F + 4(\|S^{-1}\|_F\|\Delta S\|_F)^2. \tag{21}$$

Proof. Since Y is of full rank, S is nonsingular. Thus, the orthogonality conditions can be expressed as

$$Y^T Y = S^{-1} + S^{-T} \tag{22}$$

and hence

$$\|Y^{\mathrm{T}}Y\| \leq 2\|S^{-1}\|,$$

where the norm is either the Frobenius norm or the 2-norm. Let $\Delta Q = \tilde{Q} - Q = Y\Delta S Y^{\mathrm{T}}$. Then

$$\begin{aligned}
\|\tilde{Q}^{\mathrm{T}}\tilde{Q} - I\|_F &= \|Q^{\mathrm{T}}\Delta Q + (\Delta Q)^{\mathrm{T}}Q + (\Delta Q)^{\mathrm{T}}(\Delta Q)\|_F \\
&\leq 2\|Y\|_2^2\|\Delta S\|_F + +\|Y(\Delta S)^{\mathrm{T}}Y^{\mathrm{T}}Y(\Delta S)Y^{\mathrm{T}}\|_F \\
&\leq 2\|Y^{\mathrm{T}}Y\|_F\|\Delta S\|_F + \|Y^{\mathrm{T}}Y\|_F\|Y^{\mathrm{T}}Y\|_2\|\Delta S\|_F^2 \\
&\leq 4\|S^{-1}\|_F\|\Delta S\|_F + 4(\|S^{-1}\|_F\|\Delta S\|_F)^2.
\end{aligned}$$

■

Hence, the "orthogonal" matrix $\tilde{Q}$ represented by Y and $\tilde{S}$ may in fact be far from being orthogonal if S is illconditioned, unless ΔS is sufficiently small. However, we are likely to incur a sizable ΔS when S is illconditioned, since the equation (22) says that, in essence, S^{-1} is computed from Y, and hence an inversion of S^{-1} has to be performed somewhere along the way to obtain S. If S is illconditioned with respect to inversion, then ΔS could be quite big [15].

The simplest and best-conditioned kernel is a multiple of the identity matrix. We know from the orthogonality condition (22) that a nonsingular symmetric kernel S is $2I$ if and only if the associated Householder basis Y is orthonormal. The transformation is then a block reflector with normalized basis and kernel. Note that even for the orthogonal elimination problem, where G is assumed to be orthonormal, we still need to do extra work and make Y orthonormal again to obtain $S = 2I$. Therefore, achieving a kernel with unity condition number is computationally too expensive under normal circumstances. We also note that Example 8 shows that there exist representations for symmetric orthogonal matrices (or equivalently, block reflectors), with ill-conditioned kernels.

Theorem 5 in [16] showed that any orthogonal matrix can be represented with a triangular kernel. As it turns out, the condition of a triangular kernel is greatly influenced by the scaling of its corresponding basis.

Theorem 12 *Let $Q = I - YSY^{\mathrm{T}}$ be a regular basis-kernel representation of the orthogonal matrix Q, where S is lower triangular. Then,*

$$s_{ii} = 2/\|y_i\|_2^2,$$

and, for $i > j$,

$$|(S^{-1})_{ij}| < \|y_i\|_2\|y_j\|_2, \text{ and } |s_{ij}| < \frac{4}{\|y_i\|_2\|y_j\|_2}. \tag{23}$$

Proof. By the orthogonality condition (22), $s_{ii} = 2/\|y_i\|_2^2$. The bounds on the elements of S^{-1} then follow from the triangularity of S and the Cauchy-Schwartz inequality. The strict inequality is due to the independence of Y's columns. Similarly, from the formulation (7) we have $\|YSe_j\|_2 = \sqrt{2S_{jj}} = 2/\|y_j\|_2$ and $|s_{ij}| = |(YSe_i)^{\mathrm{T}}(YS_j)| < \|YSe_i\|_2\|\|YSe_j\|_2$. The strict inequality is due to the fact that YS is of full rank. ∎

Similar results to Theorem 13 hold for lower triangular kernels.

The compact WY approach for the QR factorization gives a representation with triangular kernel and a lower triangular basis. If we adopt the conventional selection rule (18) for Householder vectors, the elementwise bounds on the kernel and its inverse can be tightened considerably.

Theorem 13 *Let $WY(y_1, \cdots, y_k) = I - YSY^{\mathrm{T}}$ be the WY representation of orthogonal transformations for the QR factorization, determined by the selection rule of (18) and the scaling convention*

$$y_j(j) = 1 \qquad \textit{if} \qquad y_j \neq 0. \tag{24}$$

Then,

$$1 \leq s_{ii} \leq 2,$$

and, for $i > j$,

$$|(S^{-1})_{ij}| \leq \sqrt{2}, \quad \textit{and} \quad |s_{ij}| \leq 2 \quad . \tag{25}$$

Furthermore,

$$\|S^{-1}\|_F \leq k, \quad \textit{and} \quad \|S\|_F < k + 1. \tag{26}$$

Proof. We assume w.l.o.g. that Y has no zero columns and that S is lower triangular. The selection rule (18) makes the first nonzero element of each Y's column the "dominant" element and the scaling convention (24) makes it equal to 1. Keeping in mind that Y's elements in the strict upper triangular part are zero, we have for each y_j, $1 \leq j \leq k$,

$$e_i^{\mathrm{T}} y_j = 0, \quad i < j, \quad e_j^{\mathrm{T}} y_j = 1, \quad \text{and} \quad 1 < \|y_j\|_2 \leq \sqrt{2}.$$

Thus, the 2-norm of y_j with the dominant element removed is no greater than 1, and for $i > j$, $|(S^{-1})_{ij}| = |y_i^{\mathrm{T}} y_j| \leq \sqrt{2}$. The bound on $\|S^{-1}\|_F$ then follows. Similarly, the elements of YS in the strict upper triangular are zero and we have

$$\|(YSe_j)\|_2 = \sqrt{2s_{jj}} \quad \text{and} \quad e_j^{\mathrm{T}} YSe_j = s_{jj}.$$

It can then be proved easily that the 2-norm of the vector obtained from (YSe_j) by zeroing the dominant element $e_j^{\mathrm{T}} YSe_j$ is unity. Therefore, for $i > j$,

$$|s_{ij}| = |(YSe_i)^{\mathrm{T}}(YSe_j)| \leq \sqrt{2S_{ii}} \leq 2.$$

Furthermore, we have

$$\|S\|_F^2 \leq \sum_{i=1}^{k} i s_{ii} = \sum_{i=1}^{k} \frac{2i}{y_i^{\mathrm{T}} y_i} \leq k(k+1).$$

■

Theorem 13 points out that orthogonal matrices computed by the WY approach are numerically benign if their degree is not unreasonably high. This is the case in practice, with k typically ranging from 8 to 64.

6 Preconditioning and Postconditioning

The preceding section showed the importance of the condition of the kernel for the numerical reliability of a block orthogonal transformation. We also showed that the compact WY accumulation procedure is very reliable. This reliability comes at a price, however, as one has to process A column by column, with the jth column of A being touched $2j+2$ times in Householder updates, norm computation, and scaling. We also note that the computation of the WY representation for k Householder vectors of length m takes $2k^2(m-k/3)$ flops for the WY basis and $mk(k+1)$ for the WY kernel [9, p. 212].

In this section we discuss other, more block-oriented, ways of achieving reasonably conditioned basis-kernel representations, and also consider the issue of sparsity. For now, assume that A is an $m \times k$, $m > k$, matrix of full rank, and C is the Cholesky factor of $A^T A$.

6.1 The Canonical Approach

Once we have computed the Cholesky factor C, the canonical basis

$$Y = \begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix}$$

is readily available, and the computation of the kernel

$$S = (A_1 + C)^{\dagger} C^{-T}$$

involves only computations on $k \times k$ matrices, which is very little work compared with the usual Householder QR factorization algorithm, since typically $k \ll m$.

The canonical basis can also be much sparser than the block representation generated by the compact WY procedure, as is shown in Figure 1, which shows A, the basis Y_{wy} computed by the compact WY procedure, and the canonical basis $Y_{canonical}$. Fill-in elements are denoted with an "F". A has 24 nonzeros and Y_{wy} has 55 nonzeros while $Y_{canonical}$ only has 29 nonzeros (i.e., only 5 fill elements). When a matrix $Q = I - YSY^{\mathrm{T}}$ is applied to a $m \times n$ matrix B, $n \gg k$, the computation of $Y^T B$ is a major part of the computational expense of forming $B - YS(Y^{\mathrm{T}}B)$ and requires $\mathrm{nonz}(Y) \cdot n \leq k \cdot m \cdot n$ flops, where $\mathrm{nonz}(Y)$ is the number of nonzero elements in Y. Since $Q = I - Y_{canonical}S_{canonical}Y_{canonical}^{\mathrm{T}} = I - Y_{wy}S_{wy}Y_{wy}^{\mathrm{T}}$, the computation of $Y_{wy}^{\mathrm{T}}B$ is not only more expensive than that of $Y_{canonical}^{\mathrm{T}}B$, but may also result in more fill-ins when B is sparse, further increasing the cost of the update.

Particular cases of the canonical basis and kernel occurred in previous works on the subject of block orthogonal factorizations. Using our framework, they can all be related to particular choices of the image U in the transformed problem (16).

For example, we proved in Lemma 5 that there is always a factor U such that the canonical kernel S is nonsingular, and hence Y is of full rank. Dietrich [6] used such a choice in his work to avoid the case of a rank-deficient kernel.

Kaufman [11] used essentially diagonal orthogonal factors. Her algorithm assumed that A_1, the top $k \times k$ submatrix of A, is upper triangular, or applied an initial QR factorization to A_1 first. Kaufman was also the first one to implicitly exploit the the sparsity of A preserved in the canonical basis and to observe the stability problems arising from ill-conditioned kernels.

Except for Schreiber and Parlett [13], all previous approaches tried to avoid producing a singular kernel. Our theory shows that, instead of being a problem, the singularity of a kernel can be taken advantage of, as it allows the generation of an orthogonal transformation of lower degree. We also showed constructively under what conditions such a kernel existed.

In this section we now present some ideas on how to cheaply solve the block elimination problem (4) while ensuring numerical reliability.

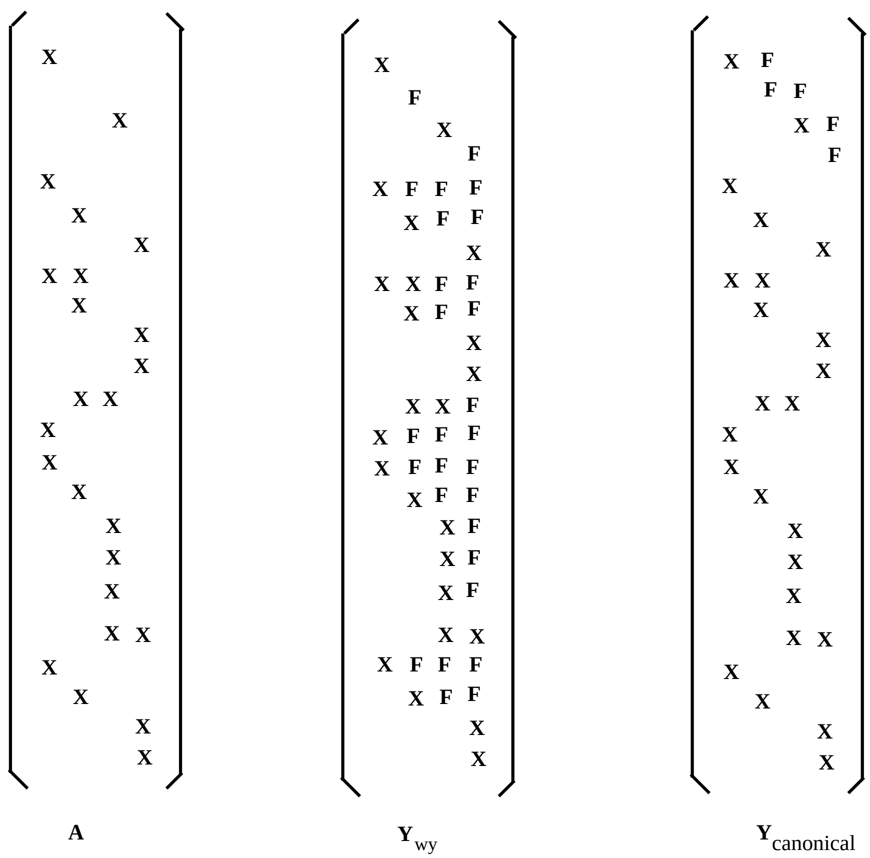

Figure 1: An example of the WY basis versus the canonical basis

6.2 Preconditioning

The canonical basis requires an admissible image of A, that is, one that satisfies the isometry condition (5). The Cholesky factor of $A^T A$ is an obvious choice. If A has orthonormal columns, the Cholesky factor is I, the identity. Therefore, the first step of the algorithms in [13] is to orthonormalize A's columns, for example, with the conventional Householder QR factorization. The conventional Householder QR procedure is also considered a reliable procedure to compute Cholesky factors. However, in either case we are essentially already solving the block elimination problem, so orthonormalizing A's columns in general is too expensive.

Computing the Cholesky $A^{\mathrm{T}}A = C^{\mathrm{T}}C$ of $A^T A$ takes $mk(k+1)$ flops for one triangular half of the matrix-matrix product $A^T A$ and $k^3/3$ flops for the Cholesky factorization. Note that this is nearly half the effort required for the WY accumulation (most of the work is in a matrix-matrix multiplication) and that the computational cost for a sparse A could even be less. We now discuss what we call "preconditioning," that is, strategies for transforming the problem so we can safely compute the Cholesky factor.

Problems in computing the Cholesky factorization arise when some of A's columns are only weakly independent of the others. This case can be dealt with through a rank-revealing Cholesky factorization. As with rank-revealing QR factorizations [5,2,3], there is a permutation matrix P such that

$$P^T(A^{\mathrm{T}}A)P = C^T C, \qquad C = \begin{pmatrix} C_{11} & C_{12} \\ & C_{22} \end{pmatrix}, \tag{27}$$

where $C_{11} \in \mathcal{R}^{r\times r}$ is wellconditioned, r is the numerical rank of A, and $\|C_{22}\|_2$ is "small."

If C_{22} is numerically negligible, then the last $k-r$ columns of AP can be considered linearly dependent on the first r ones, and we need only to determine an orthogonal matrix to solve the block elimination problem for the first r columns of AP. If R_{22} is not numerically negligible, let $\hat{R}$ be a matrix such that $C\hat{R}$ is wellconditioned, for example,

$$\hat{R} = \begin{pmatrix} I & \\ & R_{22}^{\dagger} \end{pmatrix}.$$

Now, working on the matrix $AP\hat{R}$ instead of A, we obtain

$$Q(Y,S)AP\hat{R} = \begin{pmatrix} -U\hat{C} \\ 0 \end{pmatrix}$$

for some orthogonal matrix U, and so

$$Q(Y,S)AP = \begin{pmatrix} -U\hat{C}\hat{R}^{-1} \\ 0 \end{pmatrix}.$$

Note that the Cholesky factor $\hat{C}$ of $AP\hat{R}$,

$$\hat{C} = \begin{pmatrix} C_{11} & C_{12}R_{22}^{\dagger} \\ & I \end{pmatrix},$$

is wellconditioned for the canonical basis of $AP\hat{R}$, which differs from A in only $k-r$ columns.

Note that a poor scaling in A's columns can easily result in a large condition number for $A^{\mathrm{T}}A$. In this case, the condition of A can easily be improved by scaling A's columns; that is, we choose $\hat{R}$ to be diagonal. This is probably sufficient in most cases, and certainly preferable in the sparse setting. Also note that, from the preconditioning point of view, computing an orthonormal basis of A can be considered the ultimate preconditioning step, as it results in $\hat{C} = I$.

6.3 Postconditioning

As we already noted in the context of block reflectors, a well-conditioned A does not necessarily result in a well-conditioned kernel, which is needed for numerical stability.

So now assume that we have determined a picture C of A that results in an ill-conditioned kernel, and hence an numerically rank-deficient basis. As was shown in the proof of Lemma 2 in [16], we can derive a factorization $S = F\hat{S}F^T$ of S so that $\hat{S}$ is wellconditioned, and then use the conditioned basis-kernel representation $Q(Y,S) = Q(YF,\hat{S})$. In fact, F can be composed of a permutation matrix and a triangular matrix similar to the preconditioning matrix $\hat{R}$. In general, we expect such a postconditioning matrix to drop and/or change only a small number of Y's columns. Again, the extreme case of postconditioning is the orthonormalization of Y.

Finally, note that both pre- and postconditioning involve only $k \times k$ matrices. In typical applications of block orthogonal transformations, k ranges from 4 to 32. Thus the matrices involved in conditioning are small. We have also seen from the example in Figure 1 that the canonical basis is promising for sparse orthogonal factorizations. With proper conditioning techniques, one should be able to preserve most of this desirable structure except in rare circumstances.

7 Conclusions

In this paper, we investigated the block elimination problem

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

employing the basis-kernel representation

$$Q = Q(Y, S) = I - YSY^T$$

as our main tool.

We showed that, given a particular fixed picture C, there is a unique orthogonal transformation of minimum degree, equal to $\mathrm{rank}(\begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix})$ that solves the block elimination problem. We introduced the canonical basis and kernel

$$Y = \begin{pmatrix} A_1 + C \\ A_2 \end{pmatrix} \text{ and } S = (A_1 + C)^{\dagger} C^{-T}$$

as a particularly convenient way for computing this transformation.

Considering a transformed problem where A has orthogonal columns, we then proved that, for all admissible choices of C, the minimum degree that is possible is $\mathrm{rank}(A_2)$. We showed that symmetric orthogonal matrices (i.e., block reflectors) and the compact WY representation can be considered as special cases in our general framework.

We also illustrated that the condition of the kernel S plays an important role for the numerical reliability with which $Q(Y, S)$ can be applied. We showed that an ill-scaled basis almost certainly results in a badly conditioned kernel and that the kernel computed by the compact WY accumulation procedure with the usual sign choice for Householder vectors is very wellconditioned.

Once the Cholesky factor of $A^T A$ is known, the canonical basis and kernel are much easier to compute than the compact WY accumulation strategy in the usual case where $k \ll m$. We also gave an example showing that the canonical basis and kernel hold great promise for sparse computations, since the sparsity structure of A_2 is preserved in the canonical basis. We then suggested preconditioning strategies to make sure that the Cholesky factor of $A^T A$ can be computed reliably, and postconditioning strategies to make sure that the resulting kernel is wellconditioned.

In particular for sparse problems, the canonical basis and kernel hold great promise for more efficient approaches to compute sparse orthogonal factorizations. We believe that simple column scaling strategies are sufficient as pre- and postconditioning strategies in most cases and that the investigation of pre- and postconditioning strategies for sparse matrices is a very promising avenue to pursue. Even for dense problems, we believe this approach to be worth pursuing, as the canonical basis and kernel can be computed faster than the compact WY accumulation procedure, and in a much more block-oriented fashion, which should be advantageous in cache-based systems or parallel processors. Lastly, we point out that there seems significant potential in studying how other choices for the image of A under Q can result in lower-rank and hence computationally more advantageous orthogonal transformations.

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] Christian H. Bischof and Per Christian Hansen. A block algorithm for computing rank-revealing QR factorizations. *Numerical Algorithms*, 2(3-4):371–392, 1992.

[4] Christian H. Bischof and Charles F. Van Loan. The WY representation for products of Householder matrices. *SIAM Journal on Scientific and Statistical Computing*, 8:s2–s13, 1987.

[5] Tony F. Chan. Rank revealing QR factorizations. *Linear Algebra and Its Applications*, 88/89:67–82, 1987.

[6] G. Dietrich. A new formulation of the hypermatrix Householder-QR decomposition. *Computer Methods in Applied Mechanical Engineering.*, 9:273–280, 1976.

[7] Jack Dongarra and Sven Hammarling. *Evolution of Numerical Software for Dense Linear Algebra*, pages 297–327. Oxford University Press, Oxford, UK, 1989.

[8] Jack J. Dongarra, Iain S. Duff, Danny C. Sorensen, and Henk A. Van der Vorst. *Solving Linear Systems on Vector and Shared-Memory Computers*. SIAM, Philadelphia, 1991.

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

[10] Nicholas J. Higham. Computing the polar decomposition with applications. *SIAM Journal on Scientific and Statistical Computing*, 7:1160–1174, 1986.

[11] Linda Kaufman. The generalized Householder transformation and sparse matrices. *Linear Algebra and Its Applications*, 90:221–234, 1987.

[12] Beresford N. Parlett. Analysis of algorithms for reflections in bisectors. *SIAM Review*, 13:197–208, 1971.

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

[14] 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.

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

[16] Xiaobai Sun and Christian Bischof. A basis-kernel representation of orthogonal matrices. Preprint MCS-P431-0594, Mathematics and Computer Science Division, Argonne National Laboratory, 1994.