SIAM J. MATRIX ANAL. APPL.
Vol. 40, No. 4, pp. 1257–1290

BLOCK MODIFIED GRAM–SCHMIDT ALGORITHMS AND THEIR ANALYSIS*

JESSE L. BARLOW†

Abstract. New block modified Gram–Schmidt (BMGS) methods for the Q-R factorization of a full column rank matrix $X \in \mathbf{R}^{m\times n}$, $m \geq n$, are considered. Such methods factor X into $Q \in \mathbf{R}^{m\times n}$ and an upper triangular $R \in \mathbf{R}^{m\times n}$ such that $X = QR$, where, in exact arithmetic, Q is left orthogonal (i.e., $Q^T\, Q = I_n$). Gram–Schmidt-based algorithms play an important role in the implementation of Krylov space methods, such as GMRES, Arnoldi, and Lanczos. For these applications, the left orthogonal factor Q is needed and the matrix is produced either one column at a time or one block of columns at a time. For block implementations of Krylov methods, a block of columns of X is introduced at each step, and a new block of columns of Q must then be produced. That is a task for which BMGS methods are ideally suited. However, for these Krylov methods to converge properly, the BMGS algorithms need to have numerical behavior that is similar to that of modified Gram–Schmidt. To design such BMGS algorithms, we build upon the block Householder representation of Schreiber and Van Loan [*SIAM J. Sci. Stat. Comput.*, 10 (1989), pp. 53–57] and an observation by Charles Sheffield analyzed by Paige [*SIAM J. Matrix Anal. Appl.,* 31 (2009), pp. 565–583] about the relationship between modified Gram–Schmidt and Householder Q-R factorization. Our new BMGS algorithms exploit the Sheffield framework so that they share a similar relationship to Householder Q-R and thus have error analysis properties similar to modified Gram–Schmidt. The last BMGS algorithm developed is based entirely upon matrix multiplications and the "tall, skinny" Q-R (TSQR) factorization—two operations that have been studied extensively for cache-based architectures and distributed architectures. It is shown that if the TSQR part of the BMGS algorithm satisfies error analysis properties connected to the Sheffield structure, then so does the entire BMGS algorithm. Thus new criteria for when a BMGS algorithm has error analysis properties similar to those of MGS are proposed.

Key words. Q-R decompositions, matrix-matrix operations, block algorithms, orthogonality, residuals, error bounds

AMS subject classifications. 65F25, 65F20, 65F35

DOI. 10.1137/18M1197400

1. Introduction.

For given integers m and n, $m \geq n$, and a full column rank matrix $X \in \mathbf{R}^{m\times n}$, we develop block modified Gram–Schmidt (BMGS) algorithms that factor X into $Q \in \mathbf{R}^{m\times n}$ and upper triangular, nonsingular $R \in \mathbf{R}^{n\times n}$ such that

$$X = QR \tag{1.1}$$

where, in exact arithmetic, Q is *left orthogonal*, i.e., $Q^T\, Q = I_n$.

Gram–Schmidt-based algorithms are important for the implementation of Krylov space methods, such as GMRES [29]. Analogously, BMGS algorithms are appropriate for implementing block GMRES methods that solve linear systems with multiple right-hand sides (cf. Soodhalter [35], Simoncini and Gallopoulos [34, 33]); they can also be used in implementing the block Lanczos algorithm of Golub and Underwood [21], and in the implementation of communication-avoiding Lanczos algorithms discussed by Hoemmen [24], Gustafsson, Demmel, and Holgren [22], Yamazaki and Wu [42], and Carson and Demmel [13]. Some of these works have used classical Gram–Schmidt [22, 42]; we show that BMGS is a reasonable option.

*Received by the editors June 29, 2018; accepted for publication (in revised form) June 7, 2019; published electronically October 29, 2019.
https://doi.org/10.1137/18M1197400

†Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA 16802 (barlow@cse.psu.edu).

The primary reasons for choosing Gram–Schmidt algorithms for implementing Krylov space methods are (1) the matrix X is presented to use one column at a time or one block column at a time; (2) the left orthogonal factor Q is needed; (3) a form of backward stability is necessary for the Krylov space methods to converge properly (cf. Walker [40], Paige, Rozložník, and Strakoš [29]). Thus BMGS algorithms are a good choice for implementing block Krylov methods provided that they can be shown to have numerical properties similar to those of modified Gram–Schmidt.

Our BMGS algorithms are built upon rewriting the modified Gram–Schmidt (MGS) [19, section 5.2.8] Q-R factorization as a block or BLAS-3 algorithm [16], that is, one primarily based upon matrix-matrix operations.

The framework for BMGS blends two closely related ideas. The first idea is a connection between Householder and MGS Q-R factorization first observed by Charles Sheffield and communicated to Gene Golub. That observation states that the matrices Q and R from MGS implicitly produce the Q-R factorization

$$\overline{X} \stackrel{\text{def}}{=} \begin{pmatrix} 0_{n\times n} \\ X \end{pmatrix} = U \begin{pmatrix} R \\ 0_{m\times n} \end{pmatrix}, \tag{1.2}$$

where $U \in \mathbf{R}^{(m+n)\times(m+n)}$ is a product of Householder matrices and therefore near left orthogonal in floating point arithmetic. Moreover, this equivalence holds in finite precision arithmetic. We design BMGS algorithms that are able to maintain this structure with small backward error in floating point arithmetic.

The Sheffield structure has been used by Björck and Paige [11] and by Paige, Rozložník, and Strakoš [29] in the development and analysis of MGS related algorithms. It has also been used by Barlow, Bosner, and Drmač [5], by Bosner and Barlow [12], and by Barlow [3] in the development of bidiagonalization algorithms. Paige [28] gives an exposition of the theory behind Sheffield's observation.

The second idea is the Schreiber–Van Loan [32] representation of products of Householder transformations. Since the columnwise version of MGS [19, section 5.2.8] is a BLAS-1 [27] (vector operation–based) algorithm, this representation is useful in developing matrix-vector–based (BLAS-2) and block (BLAS-3) versions of MGS. We construct a block version that uses a "tall, skinny Q-R" (TSQR) factorization [20, 24, 2] other than MGS for a key intermediate computation. This structure allows us to give sufficient conditions for a BMGS algorithm to have properties in floating point arithmetic similar to those of MGS.

The main contribution of this work is to show how the Schreiber–Van Loan representation [32] helps us build BMGS algorithms and how the use of the Sheffield structure implies that resulting BMGS algorithms inherit a key MGS property that leads to a backward stable factorization of the form (1.2) with a near left orthogonal U in floating point arithmetic. It is shown that if the TSQR part of the BMGS algorithm satisfies error analysis properties similar to MGS, so does the entire BMGS algorithm. The final proposed BMGS algorithm, called BMGS_H, is implemented with just matrix multiplications and TSQR factorizations. The fast implementation of matrix multiplications [17, 15] and of TSQR factorizations [2, 14] has been studied in a number of distributed computing environments.

Previous BMGS algorithms presented by Jalby and Philippe [26] and Vanderstraeten [39] can be understood within the structure presented here. In [26], there is a penalty in loss of orthogonality in Q in going to block algorithms; in [39] that penalty is alleviated by refactoring the diagonal blocks. Our new algorithms avoid both the penalty in loss of orthogonality in Q and the refactoring. The BCGS2 algorithm is by Barlow and Smoktunowicz [6]. It is also built upon matrix multiplications

and TSQR operations—the two key building blocks of BMGS algorithms; however, our final version of the BMGS algorithm (Function 3.3) requires one fewer TSQR operation per step than BCGS2.

Modified Gram–Schmidt is discussed by Rice [31] in an important experimental paper. The theory behind the algorithm's error analysis properties was first laid down by Björck [8]. The block classical Gram–Schmidt algorithm with reorthogonalization (BCGS2) is given and analyzed by Barlow and Smoktunowicz [6] and is a generalization of the reorthogonalized classical Gram–Schmidt algorithms of Abdelmalek [1] and Giraud et al. [18]. BCGS2 is used to develop a block downdating algorithm in [4]. A similar block algorithm based upon classical Gram–Schmidt (CGS), justified only by numerical tests, is proposed by Stewart [38]. Other interesting and closely related block orthogonal decompositions are discussed by Stathopoulos and Wu [36] and Yamazaki, Tomov, and Dongarra [41]. For a summary of the role of Gram–Schmidt algorithms, see [10, sections 2.4 and 3.2] or [9].

In the next section, we set up the remainder of the paper by reviewing the MGS algorithm and relating the Sheffield structure to the Schreiber–Van Loan Householder Q-R representation. In section 3, we give three algorithms with MGS-like properties. First, in section 3.1, we simply rewrite the MGS algorithm into matrix-vector–based and block algorithms using the ideas in section 2, whereas in section 3.2 we replace a key intermediate Q-R decomposition step that uses MGS with one that uses Householder factorization.

The error analysis of the three algorithms in section 3 is presented in section 4. That analysis begins with an appropriate perturbation theory in section 4.1. The matrix-vector–based algorithm given in section 3.1 is analyzed in section 4.2. The new BMGS algorithms in sections 3.1 and 3.2 are analyzed in section 4.3. After bounding three key residuals, all of the algorithms in section 3 are shown to fit the Sheffield structure outlined in section 2.3 with small backward error. Moreover, we give a set of conditions for which BMGS algorithms can be shown to conform to that structure with small backward error. The analysis in section 4.3 assumes that X is not so ill-conditioned as to be considered "singular to machine precision."

In section 5, we review an earlier BMGS algorithm of Jalby and Philippe [26], along with a proposed modification by Vanderstraeten [39], and interpret them in the context of our work in sections 3 and 4. The proofs of two key theorems from section 4.3 are given in Appendix A. A conclusion is given in section 6.

Five MGS-like algorithms are given as MATLAB-like functions below. They are MGS (Function 2.1), MGS2 (Function 3.1), MGS3 (Function 3.2), BMGS_H (Function 3.3), and BMGS_JP (Function 5.1). In the text below, we often refer to these functions by their abbreviations, e.g., MGS2 instead of Function 3.1.

2. MGS, the Sheffield connection, and block Householder transformations. To set the stage for the BMGS algorithms in section 3, we introduce block notation in section 2.1, we review the columnwise MGS algorithm in section 2.2, and we produce the block Householder structure of MGS in section 2.3. The Sheffield connection between MGS and Householder given in section 2.3 has been known since the 1960s [28]. What is new in this section is that we express this connection in terms of the Schreiber–Van Loan block Householder transformations [32]. That expression is used in section 3 to build the BMGS algorithms. The Sheffield connection is used to construct the error analysis in section 4.

2.1. Some block notation. To build the MGS-type algorithms, we partition X and Q into

$$X = (X_1, X_2, \ldots, X_s), \quad Q = (Q_1, Q_2, \ldots, Q_s), \tag{2.1}$$

where $X_k, Q_k \in \mathbf{R}^{m\times p_k}, k = 1, \ldots, s$, and $n = \sum_{k=1}^{s} p_k$. For the sake of simplicity, our block algorithms and their analysis assume that $p_1 = p_2 = \cdots = p_s = p$ and $n = ps$. The analysis would hold with $p = \max_{1\le k\le s} p_k$.

We denote the $m \times (kp)$ matrices made up of the first k blocks of X and Q above as

$$\widehat{X}_k = (X_1, X_2, \ldots, X_k), \quad \widehat{Q}_k = (Q_1, Q_2, \ldots, Q_k). \tag{2.2}$$

We also use this notation for the special cases when $p = 1$ with $X_k = \mathbf{x}_k$ and $Q_k = \mathbf{q}_k$ for $\mathbf{x}_k, \mathbf{q}_k \in \mathbb{R}^m$. Note that throughout this paper, the integer k *indexes the number of blocks*; it indexes the number of columns only when $p = 1$.

2.2. The MGS algorithm. The columnwise version of the MGS algorithm [10, p. 62] applied to $X \in \mathbf{R}^{m\times n}$ is given next. The algorithm as stated below is entirely based upon vector operations (i.e., BLAS-1 [27]).

FUNCTION 2.1 (Modified Gram–Schmidt).
function $[Q, R] = \mathrm{MGS}(X)$
(1) $[m, n] =$ **size**(X)*;*
(2) $r_{11} = \|\mathbf{x}_1\|_2$*;* $\mathbf{q}_1 = \mathbf{x}_1/r_{11}$*;*
(3) **for** $k = 2\colon n$
(4) $\quad \mathbf{y}_k = \mathbf{x}_k$*;*
(5) $\quad$ **for** $j = 1\colon k-1$
(6) $\quad\quad r_{jk} = \mathbf{q}_j^T \mathbf{y}_k$*;*
(7) $\quad\quad \mathbf{y}_k = \mathbf{y}_k - r_{jk}\mathbf{q}_j$;
(8) $\quad$ **end***;*
(9) $\quad r_{kk} = \|\mathbf{y}_k\|_2$*;* $\mathbf{q}_k = \mathbf{y}_k/r_{kk}$*;*
(10) **end***;*
$R = (r_{jk})$*;* $Q = (\mathbf{q}_1, \ldots, \mathbf{q}_n)$*;*
end. MGS

2.3. Block Householder structure of MGS. Our first step in constructing block versions of MGS is to rewrite the factorization (1.2) by combining the block Householder representation given by Schreiber and Van Loan [32] with a framework discovered by Charles Sheffield.

To review the Schreiber–Van Loan representation [32], for the integers m and n in (1.1), consider an orthogonal matrix $U \in \mathbf{R}^{(m+n)\times(m+n)}$ given by

$$U = P_1 \cdots P_n, \tag{2.3}$$

where, for $k = 1, \ldots, n$,

$$P_k = I_{m+n} - \mathbf{w}_k \mathbf{w}_k^T, \quad \|\mathbf{w}_k\|_2 = \sqrt{2}, \tag{2.4}$$

is a Householder transformation (cf. [25], [19, section 5.1.2]).

To represent U in (2.3)–(2.4), Schreiber and Van Loan begin with the matrix $W \in \mathbf{R}^{m\times n}$ of the Householder vectors in (2.4) given by

$$W = (\mathbf{w}_1, \ldots, \mathbf{w}_n). \tag{2.5}$$

They then show that U may be represented as

$$U = I_{m+n} - WTW^T \tag{2.6}$$

for a unit upper triangular T. Letting

$$W_k = W(:, 1\colon k) = (\mathbf{w}_1, \ldots, \mathbf{w}_k), \quad T_k = T(1\colon k, 1\colon k),$$

Schreiber and Van Loan [32] produce T from the recurrence

$$T_1 = T(1,1) = (1), \tag{2.7}$$

$$T_k = \begin{array}{c} \\ k-1 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} k-1 & 1 \end{array} \\ \begin{pmatrix} T_{k-1} & \mathbf{g}_k \\ 0 & 1 \end{pmatrix} \end{array}, \tag{2.8}$$

$$\mathbf{g}_k = -T_{k-1} W_{k-1}^T \mathbf{w}_k. \tag{2.9}$$

Puglisi [30] points out that if we let $S \in \mathbf{R}^{n\times n}$ be the unit upper triangular matrix such that

$$W^T W = S + S^T, \tag{2.10}$$

then, in exact arithmetic,

$$T = S^{-1}. \tag{2.11}$$

As Charles Sheffield pointed out to Gene Golub, a particular form of U is associated with MGS. If MGS is used to produce the Q-R factorization in (1.1), the resulting matrices Q and R satisfy (1.2), where U is given by (2.3) and P_k is the Householder transformation from (2.4) with

$$\mathbf{w}_k = \begin{pmatrix} -\mathbf{e}_k \\ \mathbf{q}_k \end{pmatrix}, \quad k = 1, \ldots, n. \tag{2.12}$$

For the $\mathbf{w}_k$ in (2.12), the matrix W from (2.13) is

$$W = \begin{pmatrix} -I_n \\ Q \end{pmatrix}. \tag{2.13}$$

The new points we wish to make in this section connect the Schreiber–Van Loan form with Sheffield's insight. Namely, some algebra applied to the expression (2.10) reveals that, for W in (2.13), S is given by

$$S = \mathbf{triu}(Q^T Q), \tag{2.14}$$

where $\mathbf{triu}(\cdot)$ denotes the upper triangular part of the contents.

In the recurrence (2.7)–(2.9), the first statement in (2.9) becomes

$$\mathbf{g}_k = -T_{k-1} \widehat{Q}_{k-1}^T \mathbf{q}_k, \tag{2.15}$$

where $\widehat{Q}_{k-1} = (\mathbf{q}_1, \ldots, \mathbf{q}_{k-1})$ results from the first $k-1$ steps of MGS. Moreover, we have that $S_k = S(1\colon k, 1\colon k)$ and (in exact arithmetic) T_k in (2.7)–(2.9) satisfy

$$S_k = \mathbf{triu}(\widehat{Q}_k^T \widehat{Q}_k), \quad T_k = S_k^{-1}. \tag{2.16}$$

Using the definition of W in (2.13), the following derivation yields a block form for U:

$$\begin{aligned} U &= I_{m+n} - WTW^T \\ &= I_{m+n} - \begin{pmatrix} -I_n \\ Q \end{pmatrix} T \begin{pmatrix} -I_n & Q^T \end{pmatrix} \\ &\qquad\qquad\quad \begin{matrix} n & \quad m \end{matrix} \\ &= \begin{matrix} n \\ m \end{matrix} \begin{pmatrix} I_n - T & TQ^T \\ QT & I_m - QTQ^T \end{pmatrix}. \end{aligned} \tag{2.17}$$

The structure in (2.17) is given in Paige's [28] discussion of Sheffield's insight.

The matrix $\widetilde{Q} \in \mathbf{R}^{m\times n}$ given by

$$\widetilde{Q} = QT = U(n+1\colon m+n, 1\colon n) \tag{2.18}$$

from the lower left block of U plays an important role in subsequent discussions.

A key point in our discussions below is that since the matrix U is the product of Householder transformations, in finite precision arithmetic with machine unit ε_M, even when the computed Q is far from left orthogonal, U will maintain near orthogonality in the sense that $\|U - \widetilde{U}\|_F = \mathcal{O}(\varepsilon_M)$ for some exactly orthogonal matrix $\widetilde{U}$. That allows us to develop algorithms in the next section that yield conditionally backward stable factorizations of the form (1.2).

3. Matrix-vector and block generalizations of MGS. In section 3.1, we use the connections outlined in section 2.3 to develop MGS2, a matrix-vector–based version of MGS, and then use the block structure in section 2.1 to build MGS3, a block version of MGS. In section 3.2, we develop BMGS_H, a BMGS algorithm based solely upon matrix multiply, and add operations and the TSQR factorization.

3.1. Using the Sheffield structure and developing a block algorithm. In order to construct a block version of MGS, we first construct a matrix-vector–based version and then generalize that to blocks. For a given X, to reformulate MGS into a matrix-vector–based function, we use (2.3), the definition of P_k in (2.4), and $\mathbf{w}_k$ in (2.12) to rewrite Function 2.1 in terms of the factorization (1.2). The k*th* step may be constructed from

$$U_{k-1} = P_1 \cdots P_{k-1},$$

which has the form

$$U_{k-1} = \begin{pmatrix} I_{k-1} - T_{k-1} & 0 & T_{k-1}\widehat{Q}_{k-1}^T \\ 0 & I_{n-k+1} & 0 \\ \widehat{Q}_{k-1}T_{k-1} & 0 & I_m - \widehat{Q}_{k-1}T_{k-1}\widehat{Q}_{k-1}^T \end{pmatrix},$$

where $\widehat{Q}_{k-1} = (\mathbf{q}_1, \ldots, \mathbf{q}_{k-1})$ is from MGS.

The *kth* columns of Q and R are computed according to

$$\begin{aligned} P_{k-1}\cdots P_1 \begin{pmatrix} 0 \\ \mathbf{x}_k \end{pmatrix} &= U_{k-1}^T \begin{pmatrix} 0 \\ \mathbf{x}_k \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{h}_k \\ \mathbf{x}_k - \widehat{Q}_{k-1}\mathbf{h}_k \end{pmatrix}, \end{aligned}$$

where

$$\mathbf{h}_k = T_{k-1}^T \widehat{Q}_{k-1}^T \mathbf{x}_k, \quad \mathbf{y}_k = \mathbf{x}_k - \widehat{Q}_{k-1}\mathbf{h}_k.$$

As done in MGS, we compute $r_{kk} = \|\mathbf{y}_k\|_2$ and $\mathbf{q}_k = \mathbf{y}_k / r_{kk}$.

Combining these operations with the recurrence for T_k given in (2.7)–(2.8), we have MGS2 expressed as Function 3.1 which also outputs T.

FUNCTION 3.1 (matrix-vector–based MGS algorithm).
function $[Q, R, T] = \text{MGS2}(X)$
(1) $[m, n] = \mathbf{size}(X)$
(2) $r_{11} = \|\mathbf{x}_1\|_2$; $\mathbf{q}_1 = \mathbf{x}_1 / r_{11}$;
(3) $R = (r_{11})$; $\widehat{Q}_1 = (\mathbf{q}_1)$, $T_1 = (1)$;
(4) **for** $k = 2\colon n$
(5) $\quad \mathbf{h}_k = \widehat{Q}_{k-1}^T \mathbf{x}_k$;
(6) $\quad \mathbf{h}_k = T_{k-1}^T \mathbf{h}_k$;
(7) $\quad \mathbf{y}_k = \mathbf{x}_k - \widehat{Q}_{k-1}\mathbf{h}_k$;
(8) $\quad r_{kk} = \|\mathbf{y}_k\|_2$; $\mathbf{q}_k = \mathbf{y}_k / r_{kk}$;
(9) $\quad \mathbf{g}_k = \widehat{Q}_{k-1}^T \mathbf{q}_k$;
(10) $\quad \mathbf{g}_k = -T_{k-1}\mathbf{g}_k$;
(11) $\quad \widehat{Q}_k = \begin{pmatrix} \widehat{Q}_{k-1} & \mathbf{q}_k \end{pmatrix}$; $R_k = \begin{pmatrix} R_{k-1} & \mathbf{h}_k \\ 0 & r_{kk} \end{pmatrix}$;
(12) $\quad T_k = \begin{pmatrix} T_{k-1} & \mathbf{g}_k \\ 0 & 1 \end{pmatrix}$;
(13) **end**;
$Q = \widehat{Q}_n$; $R = R_n$; $T = T_n$;
end. MGS2

MGS2 gives us Q and R satisfying (1.1) and T satisfying (2.10)–(2.11).

The important operations in MGS2 are (5), (6), (7), (9), and (10), and these are all matrix-vector products.

To build a block version of MGS, we assume that X and Q are partitioned into s blocks of size $m \times p$ as in (2.1) and let R be partitioned according to

$$R = \begin{pmatrix} R_{11} & R_{12} & \cdots & \cdots & \cdots & R_{1s} \\ 0 & R_{22} & R_{23} & \cdots & \cdots & R_{2s} \\ 0 & 0 & R_{33} & \cdots & \cdots & R_{3s} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & R_{s-1,s-1} & R_{s-1,s} \\ 0 & \cdots & \cdots & \cdots & 0 & R_{ss} \end{pmatrix}, \quad R_{ij} \in \mathbf{R}^{p \times p}, \tag{3.1}$$

and we assume that T and S in (2.14) are partitioned conformally. We let R_k be given by

$$R_k = \begin{pmatrix} R_{11} & R_{12} & \cdots & \cdots & \cdots & R_{1k} \\ 0 & R_{22} & R_{23} & \cdots & \cdots & R_{2k} \\ 0 & 0 & R_{33} & \cdots & \cdots & R_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & R_{k-1,k-1} & R_{k-1,k} \\ 0 & \cdots & \cdots & \cdots & 0 & R_{kk} \end{pmatrix} \tag{3.2}$$

and let

$$H_k = \begin{pmatrix} R_{1k} \\ R_{2k} \\ \vdots \\ R_{k-1,k} \end{pmatrix}. \tag{3.3}$$

Also, we define T_k conformally to R_k in (3.3) and let G_k be given by

$$G_k = \begin{pmatrix} T_{1k} \\ T_{2k} \\ \vdots \\ T_{k-1,k} \end{pmatrix}. \tag{3.4}$$

We can then block partition the MGS2 version of MGS as follows:

$$[Q_1, R_{11}, T_{11}] = \mathbf{MGS2}(X_1), \tag{3.5}$$

$$H_k = \widehat{T}_{k-1}^T \widehat{Q}_{k-1}^T X_k, \quad k = 1, \ldots, s, \tag{3.6}$$

$$Y_k = X_k - \widehat{Q}_{k-1} H_k, \tag{3.7}$$

$$[Q_k, R_{kk}, T_{kk}] = \mathbf{MGS2}(Y_k). \tag{3.8}$$

To obtain a formula for G_k in (3.4) while maintaining the relationship (2.16), we note that the (k, k) block of S in (2.10) is given by

$$S_{kk} = \mathbf{triu}(Q_k^T Q_k), \tag{3.9}$$

we use the standard block inverse formula

$$S_k^{-1} = \begin{pmatrix} S_{k-1}^{-1} & -S_{k-1}^{-1} \widehat{Q}_{k-1}^T Q_k S_{kk}^{-1} \\ 0 & S_{kk}^{-1} \end{pmatrix}, \tag{3.10}$$

and we enforce $T_k = S_k^{-1}$ to yield the recurrence

$$T_k = \begin{pmatrix} T_{k-1} & G_k \\ 0 & T_{kk} \end{pmatrix} \tag{3.11}$$

where

$$G_k = -T_{k-1} \widehat{Q}_{k-1}^T Q_k T_{kk}. \tag{3.12}$$

Thus, we have the factorization

$$\widehat{X}_k = \widehat{Q}_k R_k, \quad \widehat{Q}_k = \begin{pmatrix} \widehat{Q}_{k-1} & Q_k \end{pmatrix},$$
$$R_k = \begin{pmatrix} R_{k-1} & H_k \\ 0 & R_{kk} \end{pmatrix}.$$

Using the above development, the following is a formal statement of our algorithm.

FUNCTION 3.2 (MGS3–block version of the MGS algorithm).
function $[Q, R, T]$ =MGS3(X, p)
(1) $[m, n] =$ **size**(X);$s = n/p$;
(2) $[\widehat{Q}_1, R_{11}, T_1] =$ **MGS2**(X_1);
(3) **for** $k = 2\colon s$
(4) $\quad H_k = \widehat{Q}_{k-1}^T X_k$;
(5) $\quad H_k = T_{k-1}^T H_k$;
(6) $\quad Y_k = X_k - \widehat{Q}_{k-1} H_k$;
(7) $\quad [Q_k, R_{kk}, T_{kk}] =$ **MGS2**(Y_k);
(8) $\quad F_k = \widehat{Q}_{k-1}^T Q_k$;
(9) $\quad G_k = -T_{k-1} F_k T_{kk}$;
(10) $\quad \widehat{Q}_k = \begin{pmatrix} \widehat{Q} & Q_k \end{pmatrix}$. $R_k = \begin{pmatrix} R_{k-1} & H_k \\ 0 & R_{kk} \end{pmatrix}$;
(11) $\quad T_k = \begin{pmatrix} T_{k-1} & G_k \\ 0 & T_{kk} \end{pmatrix}$;
(12) **end**;
$Q = \widehat{Q}_s$; $R = R_s$, $T = T_s$;
end. MGS3

In the next section, we modify MGS3 by substituting Householder Q-R factorization in steps (2) and (7) for MGS2.

3.2. Block MGS with Householder-based Q-R factorization. Steps (2) and (7) of MGS3 compute the factorization of Y_k using MGS2. Although using MGS2 makes MGS3 a pure Gram–Schmidt algorithm, it is a matrix-vector–based algorithm, and there are other possible choices for those two steps. Since X_1 and Y_k are usually "tall, skinny" matrices, Householder-based "tall, skinny" Q-R (TSQR) factorization is a good choice. Communication-avoiding versions of Householder-based TSQR for distributed computing have been discussed by a number of authors; see, for example, [2, 14]. Discussion of TSQR algorithms in the literature dates back at least to a 1988 paper by Golub, Plemmons, and Sameh [20]. The other major component of the algorithm below is matrix multiply and add, i.e., the BLAS-3 xGEMM operation [16], and versions of this algorithm for distributed computing environments have been discussed by a number of authors, (cf. [17, 15]).

We let the statement

$$[Q_k, R_{kk}] = \textbf{House\_QR}(Y_k) \tag{3.13}$$

denote the result of a Householder-based Q-R factorization applied to Y_k such as those in [2, 14]. Instead of producing T_{kk} as MGS2 does, we set $T_{kk} = I_p$; we show in section 4.3 that this is reasonable in floating point arithmetic. We now present a block MGS algorithm called BMGS_H that uses Householder Q-R to factor Y_k.

FUNCTION 3.3 (BMGS with Householder-based Q-R).
function $[Q, R, T]$ =BMGS_H(X, p)
(1) $[m, n] =$ **size**(X);$s = n/p$;
(2) $[\widehat{Q}_1, R_{11}] =$ **House_QR**(X_1); $T_1 = I_p$;
(3) **for** $k = 2\colon s$
(4) $\quad H_k = \widehat{Q}_{k-1}^T X_k$;
(5) $\quad H_k = T_{k-1}^T H_k$;
(6) $\quad Y_k = X_k - \widehat{Q}_{k-1} H_k$;

(7) $[Q_k, R_{kk}] = \mathbf{House\_QR}(Y_k)$;
(8) $F_k = \widehat{Q}_{k-1}^T Q_k$;
(9) $G_k = -T_{k-1}F_k$ % *Note that* $T_{kk} = I_p$;
(10) $\widehat{Q}_k = \begin{pmatrix} \widehat{Q}_{k-1} & Q_k \end{pmatrix}$. $R_k = \begin{pmatrix} R_{k-1} & H_k \\ 0 & R_{kk} \end{pmatrix}$;
(11) $T_k = \begin{pmatrix} T_{k-1} & G_k \\ 0 & I_p \end{pmatrix}$;
(12) **end**;
$Q = \widehat{Q}_s$; $R = R_s$, $T = T_s$;
end. BMGS_H

Function 3.3 (BMGS_H) is similar in structure and data movement to the block classical Gram–Schmidt algorithm BCGS2 by Barlow and Smoktunowicz [6]. However, BCGS2 does four matrix multiplications with $\widehat{Q}_{k-1}$ at each step compared with three for BMGS_H, and BGCS2 does two calls to House_QR per step compared to one for BMGS_H. BMGS_H also requires two multiplications with T_{k-1} per step, but those will cost much less than an extra multiplication with $\widehat{Q}_{k-1}$ and an extra call to House_QR. Under the assumptions of the error analysis in [6], the resulting Q from BCGS2 is near left orthogonal, whereas the Q from BMGS_H is not in general. However, when solving least squares problems, the two algorithms have a similar guarantee of accuracy. For the case $p = 1$, there is a similar trade-off between MGS2 and the cgs2 algorithm in [18].

4. Error analysis results. To begin our error analysis results for the factorizations produced by MGS2, MGS3, and BMGS_H, we develop a perturbation theory in section 4.1. For that, we let Q, R, and T be computed quantities, let S be the exact matrix defined by (2.14), and define the three residual errors

$$TS - I_n = \Delta_{TS}, \tag{4.1}$$
$$QR - X = \delta X, \tag{4.2}$$
$$(I_n - T)R = \Gamma_{TR}. \tag{4.3}$$

We show that if we have bounds on $\|\Delta_{TS}\|_F$, $\|\delta X\|_F$, and $\|\Gamma_{TR}\|_F$, there exist $\widetilde{U} \in \mathbf{R}^{(m+n)\times(m+n)}$ exactly orthogonal and $V \in \mathbf{R}^{m\times n}$ exactly left orthogonal such that we have bounds on $\|U - \widetilde{U}\|_F$ for U in (2.17), $\|Q - V\|_F$ and $\|\widetilde{Q} - V\|_F$ for $\widetilde{Q}$ in (2.18), $\|\overline{X} - \widetilde{U}\binom{R}{0}\|_F$ for $\overline{X}$ in (1.2), and $\|X - VR\|_F$. We also establish bounds on $\|Q\|_2$, $\|S\|_2$, and $\|T\|_2$. These are similar to bounds already established for MGS by Björck [8] and Björck and Paige [11].

In our floating point error analysis in sections 4.2 and 4.3, we follow a convention in [19, section 2.7.7] by producing first order bounds in the machine unit ε_M and attaching a term of $+\mathcal{O}(\varepsilon_M^2)$ where appropriate. Also, to simplify the analysis, many of the perturbation bounds in section 4.1 are in the two-norm, but the error analysis bounds in sections 4.2 and 4.3 are in the Frobenius norm.

Bounds on $\|\Delta_{TS}\|_F$, $\|\delta X\|_F$, and $\|\Gamma_{TR}\|_F$ for MGS2 are given in section 4.2.

In section 4.3, we show that for MGS3 and BMGS_H, under restrictions on R, for modest sized functions $f_{TR}(\cdot)$, $f_X(\cdot)$, and $f_{TS}(\cdot)$, we have

$$\|\Delta_{TS}\|_F \le \varepsilon_M f_{TS}(m,n,p) + \mathcal{O}(\varepsilon_M^2), \tag{4.4}$$
$$\|\delta X\|_F \le \varepsilon_M f_X(m,n,p)\|X\|_F + \mathcal{O}(\varepsilon_M^2), \tag{4.5}$$
$$\|\Gamma_{TR}\|_F \le \varepsilon_M f_{TR}(m,n,p)\|X\|_F + \mathcal{O}(\varepsilon_M^2). \tag{4.6}$$

That the recurrences (2.7)–(2.9) and (3.11)–(3.12) produce T satisfying the bound (4.4) is unsurprising. Any reasonable Q-R decomposition algorithm should satisfy a bound such as (4.5). However, (4.6) is not necessarily satisfied by all Q-R decomposition algorithms, and thus its proof is key to our error analysis.

In section 4.4, we give numerical tests demonstrating that all of the MGS-like algorithms in section 3 have numerical properties similar to those of to MGS.

The following well-known example shows a Q-R decomposition algorithm, classical Gram–Schmidt, that does not have a bound of the form (4.6) on $\|\Gamma_{TR}\|_F$.

Example 4.1. We revisit the classic Läuchli example

$$X = \begin{pmatrix} 1 & 1 & 1 \\ \eta & 0 & 0 \\ 0 & \eta & 0 \\ 0 & 0 & \eta \end{pmatrix}, \quad \eta \in (\varepsilon_M, \sqrt{\varepsilon_M}). \tag{4.7}$$

If we assume that no rounding errors are committed after the computation $r_{11} = \mathrm{fl}(\sqrt{1+\eta^2}) = 1$, then the computed Q and R from MGS and MGS2 (Functions 2.1 and 3.1) are

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ \eta & -1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 0 & \sqrt{2/3} \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 1 & 1 \\ 0 & \sqrt{2}\eta & \eta/\sqrt{2} \\ 0 & 0 & \sqrt{3/2}\eta \end{pmatrix}. \tag{4.8}$$

The formulas for S and T are

$$S = \begin{pmatrix} 1 & -\eta/\sqrt{2} & -\eta/\sqrt{6} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & \eta/\sqrt{2} & \eta/\sqrt{6} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.9}$$

Using these formulas, we have

$$\|\Gamma_{TR}\|_F = \|(I-T)R\|_F = \sqrt{2}\eta^2 < \sqrt{2}\varepsilon_M,$$

and thus a condition such as (4.6) is satisfied.

Classical Gram–Schmidt (CGS) [10, p. 63] obtains the matrices

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ \eta & -1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 1/\sqrt{2} \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 1 & 1 \\ 0 & \sqrt{2}\eta & 0 \\ 0 & 0 & \sqrt{2}\eta \end{pmatrix},$$

$$S = \begin{pmatrix} 1 & -\eta/\sqrt{2} & -\eta/\sqrt{2} \\ 0 & 1 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & \eta/\sqrt{2} & \eta/2\sqrt{2} \\ 0 & 1 & -1/2 \\ 0 & 0 & 1 \end{pmatrix}.$$

We have that

$$\Gamma_{TR} = (I-T)R = \begin{pmatrix} 0 & \eta^2 & \eta^2/(2\sqrt{2}) \\ 0 & 0 & -\eta/\sqrt{2} \\ 0 & 0 & 0 \end{pmatrix},$$

and thus $\|\Gamma_{TR}\|_F = \eta/\sqrt{2} + \mathcal{O}(\eta^2)$, implying that CGS does not satisfy (4.6). The Q matrices in this example are given in [10, p. 67].

4.1. Perturbation theory for MGS and BMGS algorithms. Our first theorem relates the perturbation (4.1) to the distance between U in (2.17) and an orthogonal matrix $\widetilde{U}$ and establishes a backward error relationship between $\widetilde{U}$, $\overline{X}$ in (1.2) and R for the perturbations (4.1)–(4.3).

THEOREM 4.1. *Let $Q \in \mathbf{R}^{m\times n}$, $R \in \mathbf{R}^{n\times n}$, and $T \in \mathbf{R}^{n\times n}$ satisfy* (4.1)–(4.3) *with R nonsingular and $\|Q\mathbf{e}_j\|_2 = 1$ for $j = 1, \ldots, n$. Let $\overline{X}$ be given by* (1.2), *U by* (2.17), *and S by* (2.14). *Then S is nonsingular, $\|S^{-1}\|_2 \le 2$,*

$$\widetilde{U} = \begin{pmatrix} I_n - S^{-1} & S^{-1}Q^T \\ QS^{-1} & I_m - QS^{-1}Q^T \end{pmatrix} \tag{4.10}$$

is exactly orthogonal,

$$\|U - \widetilde{U}\|_F \le \sqrt{10}\|Q\|_2\|\Delta_{TS}\|_F \le \sqrt{10n}\|\Delta_{TS}\|_F, \tag{4.11}$$

and

$$\overline{X} + \delta\overline{X} = \widetilde{U}\begin{pmatrix} R \\ 0_{m\times n} \end{pmatrix}, \quad \delta\overline{X} = \begin{matrix} n \\ m \end{matrix}\begin{pmatrix} \delta\overline{X}_1 \\ \delta\overline{X}_2 \end{pmatrix}, \tag{4.12}$$

where

$$\delta\overline{X}_1 = \Gamma_{TR} + \Delta_{TS}S^{-1}R, \tag{4.13}$$

$$\delta\overline{X}_2 = \delta X - Q(\delta\overline{X}_1). \tag{4.14}$$

Thus,

$$\|\delta\overline{X}_1\|_F \le \|\Gamma_{TR}\|_F + 2\|\Delta_{TS}\|_F\|R\|_2, \tag{4.15}$$

$$\|\delta\overline{X}\|_F \le \|\delta X\|_F + \left(1 + \|Q\|_2^2\right)^{1/2}\|\delta\overline{X}_1\|_F. \tag{4.16}$$

Proof. From (2.14), S is a unit upper triangular matrix, and thus S^{-1} exists. That $\widetilde{U}$ in (2.17) is orthogonal can be verified by confirming that $\widetilde{U}^T\,\widetilde{U} = I_{m+n}$. Thus,

$$\|I_n - S^{-1}\|_2 \le \|\widetilde{U}\|_2 = 1$$

so that

$$\|S^{-1}\|_2 \le \|I_n\|_2 + \|I_n - S^{-1}\|_2 \le 2.$$

From (2.17) and (4.10), we note that

$$\begin{aligned} U - \widetilde{U} &= \begin{pmatrix} S^{-1} - T & (T - S^{-1})Q^T \\ Q(T - S^{-1}) & Q(S^{-1} - T)Q^T \end{pmatrix} \\ &= \begin{pmatrix} I_n & 0 \\ 0 & Q \end{pmatrix}\Delta_{TS}\begin{pmatrix} -S^{-1} & S^{-1}Q^T \\ S^{-1} & -S^{-1}Q^T \end{pmatrix}. \end{aligned} \tag{4.17}$$

Thus,

$$\begin{aligned} \|U - \widetilde{U}\|_F &\le \left\|\begin{pmatrix} I_n & 0 \\ 0 & Q \end{pmatrix}\right\|_2 \|\Delta_{TS}\|_F \left\|\begin{pmatrix} -S^{-1} & S^{-1}Q^T \\ S^{-1} & -S^{-1}Q^T \end{pmatrix}\right\|_2 \\ &\le \|Q\|_2 \left\|\begin{pmatrix} \|S^{-1}\|_2 & \|S^{-1}Q^T\|_2 \\ \|S^{-1}\|_2 & \|S^{-1}Q^T\|_2 \end{pmatrix}\right\|_2 \|\Delta_{TS}\|_F. \end{aligned} \tag{4.18}$$

Since $S^{-1}Q^T$ is the $(1,2)$ block of the orthogonal matrix $\widetilde{U}$, $\|S^{-1}Q^T\|_2 \le 1$, and thus (4.18) reads

$$\|U - \widetilde{U}\|_F \le \|Q\|_2 \left\| \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} \right\|_2 \|\Delta_{TS}\|_F \le \sqrt{10}\|Q\|_2 \|\Delta_{TS}\|_F, \tag{4.19}$$

which is the first inequality in (4.11). Since Q has columns that are unit vectors,

$$\|Q\|_2 \le \|Q\|_F = \sqrt{n},$$

and we have the second inequality in (4.11).

To get (4.12), we use (2.17) and (4.17) and note that

$$\begin{aligned} \widetilde{U} \begin{pmatrix} R \\ O_{m\times n} \end{pmatrix} &= U \begin{pmatrix} R \\ 0_{m\times n} \end{pmatrix} + (\widetilde{U} - U) \begin{pmatrix} R \\ 0_{m\times n} \end{pmatrix} \\ &= \begin{pmatrix} (I_n - T)R \\ QTR \end{pmatrix} + \begin{pmatrix} \Delta_{TS} S^{-1} R \\ -Q\Delta_{TS} S^{-1} R \end{pmatrix} \\ &= \begin{pmatrix} \Gamma_{TR} \\ X + \delta X - Q\Delta_{TS} R \end{pmatrix} + \begin{pmatrix} \Delta_{TS} S^{-1} R \\ -Q\Delta_{TS} S^{-1} R \end{pmatrix} \\ &= \overline{X} + \delta\overline{X}, \end{aligned}$$

where $\delta\overline{X}$ and its blocks $\overline{X}_1$ and $\delta\overline{X}_2$ satisfy (4.12)–(4.14). Standard norm bounds yield (4.16)–(4.15). ∎

To understand the importance of bounding these quantities, we give a version of a theorem from Paige [28] for the case where R is nonsingular.

THEOREM 4.2 ([28]). *For $X \in \mathbf{R}^{m\times n}$, let $\overline{X}$ be given by* (1.2)*, let $\delta\overline{X}$ be given by* (4.12) *with the partitioning given there, let $\widetilde{U} \in \mathbf{R}^{(m+n)\times(m+n)}$ be orthogonal, let $Z = \widetilde{U}(:, 1\colon n) \in \mathbf{R}^{(m+n)\times n}$, let $R \in \mathbf{R}^{n\times n}$ be nonsingular, and let $\widetilde{U}$ and R satisfy* (4.12)–(4.14)*. If Z is partitioned into*

$$Z = \begin{matrix} n \\ m \end{matrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \tag{4.20}$$

then there exists a left orthogonal matrix $V \in \mathbf{R}^{m\times n}$ such that

$$X + \delta\tilde{X} = VR, \tag{4.21}$$

$$\delta\tilde{X} = FZ_1^T(\delta\overline{X}_1) + \delta\overline{X}_2 \tag{4.22}$$

for some $F \in \mathbf{R}^{m\times n}$ where $\|F\|_2 \in [0.5, 1]$.

In the context of Theorem 4.1, Z_1 in (4.20) is

$$Z_1 = I - S^{-1} = (\delta\overline{X}_1)R^{-1}. \tag{4.23}$$

We define the useful quantity

$$\zeta \stackrel{\text{def}}{=} \|Z_1\|_2 = \|(\delta\overline{X}_1)R^{-1}\|_2. \tag{4.24}$$

The following lemma justifies an assumption about ζ.

LEMMA 4.3. *Assume the hypothesis and notation of Theorem* 4.1*, let Z be as in Theorem* 4.2*, let Z_1 and Z_2 be from* (4.20)*, and let ζ be defined by* (4.24)*. Then $\zeta < 1$ if and only if* $\operatorname{rank}(Q) = n$.

Proof. First, assume that $\operatorname{rank}(Q) = n$. Then $Z_2 = QS^{-1}$ also has full column rank. For any $\mathbf{w} \in \mathbb{R}^n$ such that $\|\mathbf{w}\|_2 = 1$, we have that

$$\|Z\mathbf{w}\|_2^2 = \|Z_1\mathbf{w}\|_2^2 + \|Z_2\mathbf{w}\|_2^2 = 1.$$

Thus,

$$\begin{aligned}\zeta^2 = \|Z_1\|_2^2 &= \max_{\|\mathbf{w}\|_2=1} \|Z_1\mathbf{w}\|_2^2 \\ &= \max_{\|\mathbf{w}\|_2=1} 1 - \|Z_2\mathbf{w}\|_2^2 = 1 - \min_{\|\mathbf{w}\|_2=1} \|Z_2\mathbf{w}\|_2^2.\end{aligned}$$

Since Z_2 has full column rank, $\sigma_n(Z_2)$ (the smallest singular value of Z_2) satisfies

$$\min_{\|\mathbf{w}\|_2=1} \|Z_2\mathbf{w}\|_2 = \sigma_n(Z_2) > 0.$$

Thus,

$$\zeta = \left(1 - \sigma_n(Z_2)^2\right)^{1/2} < 1.$$

If we assume $\zeta < 1$, every step of the argument given above can be reversed to conclude that $\operatorname{rank}(Q) = n$. Thus $\zeta < 1$ and $\operatorname{rank}(Q) = n$ are equivalent conditions. ∎

In Theorem 4.4 and Corollary 4.5, we bound the distance between the exactly left orthogonal matrix V from Theorem 4.2 and Q from (1.1) and $\widetilde{Q}$ in (2.18). We also give bounds on $\|Q\|_2$ and $\|T\|_2$ that are necessary for our error analysis proof in Appendix A.

THEOREM 4.4. *Assume the hypothesis and notation of Theorem* 4.1*, let* $Z = \widetilde{U}(:, 1\colon n) \in \mathbf{R}^{(m+n)\times n}$ *be partitioned according to* (4.20)*, let* $\widetilde{Q}$ *be as* (2.18)*, let* ζ *be as in* (4.24)*, and assume that* Q *has full column rank.*

Then $\zeta < 1$*,*

$$\|Q\|_2 \le (1 + \zeta^2)/(1 - \zeta), \tag{4.25}$$

$$\|S^{-1}\|_2 \le 1 + \zeta, \tag{4.26}$$

and, for some exactly left orthogonal matrix $V \in \mathbf{R}^{m\times n}$*,*

$$\|Q - V\|_2 \le (\zeta + \zeta^2)/(1 - \zeta), \tag{4.27}$$

$$\|\widetilde{Q} - V\|_2 \le (1 + \zeta)(1 + \zeta^2)\|\Delta_{TS}\|_2/(1 - \zeta) + \zeta^2. \tag{4.28}$$

Proof. First, we note that Z_1 satisfies (4.23), and from Lemma 4.3, $\zeta < 1$. Combining (4.2) and (4.21)–(4.22), we have that

$$\begin{aligned}(Q - V)R &= \delta X - \delta\tilde{X} \\ &= \delta X - (FZ_1^T(\delta\overline{X}_1) + \delta\overline{X}_2) \\ &= \delta X - (FZ_1^T(\delta\overline{X}_1) + \delta X - Q(\delta\overline{X}_1)) \\ &= -FZ_1^T(\delta\overline{X}_1) + Q(\delta\overline{X}_1),\end{aligned}$$

where $\|F\|_2 \in [0.5, 1]$. Since R is assumed to be nonsingular and $Z_1 = (\delta\overline{X}_1)R^{-1}$, we have

$$\begin{aligned}Q - V &= -FR^{-T}(\delta\overline{X}_1)^T\,(\delta\overline{X}_1)R^{-1} + Q(\delta\overline{X}_1)R^{-1} \\ &= -FZ_1^T\,Z_1 + VZ_1 + (Q - V)Z_1.\end{aligned}$$

Thus,

$$(Q - V)(I_n - Z_1) = -FZ_1^T Z_1 + VZ_1.$$

Since $\zeta < 1$, we have that

$$\begin{aligned}\|Q - V\|_2 &\le (\|F\|_2\|Z_1\|_2^2 + \|V\|_2\|Z_1\|_2)/(1 - \|Z_1\|_2)\\ &\le (\zeta + \zeta^2)/(1 - \zeta),\end{aligned}$$

which is (4.27). The bound on $\|Q\|_2$ follows from

$$\|Q\|_2 = \|V\|_2 + \|Q - V\|_2 \le 1 + (\zeta + \zeta^2)/(1 - \zeta) = (1 + \zeta^2)/(1 - \zeta).$$

To obtain (4.26), note that

$$\|S^{-1}\|_2 = \|I_n + I_n - S^{-1}\|_2 \le \|I_n\|_2 + \|I_n - S^{-1}\|_2 = 1 + \zeta.$$

To get (4.28), we use the relationship

$$\begin{aligned}\widetilde{Q}R &= Z_2R + (\widetilde{Q} - Z_2)R = Z_2R + Q(\Delta_{TS})S^{-1}R\\ &= X + \delta\overline{X}_2 + Q\Delta_{TS}S^{-1}R.\end{aligned}$$

Thus, we have

$$\begin{aligned}(\widetilde{Q} - V)R &= \delta\overline{X}_2 + Q(\Delta_{TS})S^{-1}R - \delta\tilde{X}\\ &= \delta\overline{X}_2 + Q(\Delta_{TS})S^{-1}R - (FZ_1^T(\delta\overline{X}_1) + \delta\overline{X}_2)\\ &= Q(\Delta_{TS})S^{-1}R - FZ_1^T(\delta\overline{X}_1),\end{aligned}$$

which yields

$$\begin{aligned}\widetilde{Q} - V &= Q(\Delta_{TS})S^{-1} - FZ_1^T(\delta\overline{X}_1)R^{-1}\\ &= Q(\Delta_{TS})S^{-1} - FZ_1^T Z_1.\end{aligned}$$

Using standard norm bounds gives us

$$\begin{aligned}\|\widetilde{Q} - V\|_2 &\le \|Q\|_2\|\Delta_{TS}\|_2\|S^{-1}\|_2 + \|F\|_2\|Z_1\|_2^2\\ &\le (1 + \zeta)(1 + \zeta^2)\|\Delta_{TS}\|_2/(1 - \zeta) + \zeta^2,\end{aligned}$$

which is (4.28). ∎

Remark 4.1. Since the distance between Q and the left orthogonal matrix V is $\mathcal{O}(\zeta)$ and the distance between $\widetilde{Q}$ in (2.18) and the same V is $\mathcal{O}(\max\{\|\Delta_{TS}\|_2, \zeta^2\})$, we expect $\widetilde{Q}$ to be more nearly left orthogonal than Q.

COROLLARY 4.5. *Assume the hypotheses and terminology of Theorems* 4.1 *and* 4.4. *Then,*

$$\|T\|_2 \le (1 + \|\Delta_{TS}\|_2)(1 + \eta), \tag{4.29}$$

$$\|I_n - S\|_2 \le \zeta/(1 - \zeta), \quad \|S\|_2 \le 1/(1 - \zeta). \tag{4.30}$$

Proof. We have that

$$T = S^{-1} + T - S^{-1} = (I_n + \Delta_{TS})S^{-1}.$$

Thus,

$$\|T\|_2 = \|(I_n + \Delta_{TS})S^{-1}\|_2 \le (1 + \|\Delta_{TS}\|_2)\|S^{-1}\|_2 \le (1 + \|\Delta_{TS}\|_2)(1 + \zeta),$$

which is (4.29).

From (4.23), $\|I_n - S^{-1}\|_2 = \|Z_1\|_2 = \zeta < 1$, and

$$\|S\|_2 = \|[I + (I - S^{-1})]^{-1}\|_2 \le 1/(1 - \|I - S^{-1}\|_2) \le 1/(1 - \zeta),$$

which is the second part of (4.30). Thus,

$$I_n - S = (S^{-1} - I_n)S$$

so that

$$\|I_n - S\|_F \le \|S\|_2\|I_n - S^{-1}\|_2 \le \|S\|_2\|Z_1\|_2 \le \zeta/(1 - \zeta),$$

which is the first part of (4.30). ∎

Motivated by (4.27), we make the assumption

$$\|Q - V\|_2 \le (\zeta + \zeta^2)/(1 - \zeta) \le 1. \tag{4.31}$$

Equation (4.31) holds if Q is remotely close to an orthogonal matrix and is equivalent to the assumption that

$$\zeta \le \sqrt{2} - 1 \approx 0.41421. \tag{4.32}$$

Using the definition (4.24), the bounds (4.6) and (4.4) on $\|\Gamma_{TR}\|_F$ and $\|\Delta_{TS}\|_F$, and the bound (4.15), we can say that

$$\zeta \le \hat{\zeta} = \varepsilon_M f_\zeta(m, n, p)\|X\|_F\|R^{-1}\|_2, \tag{4.33}$$

$$f_\zeta(m, n, p) = f_{TR}(m, n, p) + 2f_{TS}(m, n, p), \tag{4.34}$$

and thus we make the assumption that

$$\zeta \le \hat{\zeta} \le \sqrt{2} - 1. \tag{4.35}$$

The assumption (4.35) is false only if X is sufficiently ill-conditioned to be $\mathcal{O}(\varepsilon_M)$ distance from a rank deficient matrix. We also make the assumption that $\|\Delta_{TS}\|_2 \le \sqrt{2} - 1$.

Equation (4.35) and Corollary 4.5 allow us to say that

$$\|Q\|_2, \|S\|_2 \le 2, \quad \|T\|_2 \le \sqrt{2}(1 + \|\Delta_{TS}\|_2) \le 2. \tag{4.36}$$

Also, for each k in MGS2, MGS3, and BMGS_H, we can say that

$$\|\widehat{Q}_k\|_2, \|Q_k\|_2, \|S_k\|_2, \|T_k\|_2, \|T_{kk}\|_2 \le 2. \tag{4.37}$$

The bounds in (4.37) simplify our error analysis.

Equation (4.37) allows us to replace (4.11) and (4.13) with

$$\|U - \widetilde{U}\|_F \le 2\sqrt{10}\|\Delta_{TS}\|_F, \tag{4.38}$$

$$\|\delta\overline{X}\|_F \le \|\delta X\|_F + \sqrt{5}(\|\Gamma_{TR}\|_F + 2\|\Delta_{TS}\|_F\|R\|_2). \tag{4.39}$$

The following corollary establishes a bound on orthogonality of Q similar to that from [8].

COROLLARY 4.6. *Assume the hypothesis and terminology of Theorem* 4.4. *Also assume* (4.35). *Then,*

$$\|I_n - Q^T Q\|_F \le 2\hat{\zeta}/(1-\hat{\zeta}). \tag{4.40}$$

Proof. Following the argument in Corollary 4.5 to prove (4.30), and bounding the Frobenius norm instead of the two-norm, we have

$$\|I_n - S\|_F \le \|S\|_2 \|Z_1\|_F \le \hat{\zeta}/(1-\zeta) \le \hat{\zeta}/(1-\hat{\zeta}).$$

From the definition of S in (2.14), we have

$$I_n - Q^T Q = I_n - S + I_n - S^T,$$

and thus,

$$\|I_n - Q^T Q\|_F \le 2\|I_n - S\|_F \le 2\hat{\zeta}/(1-\hat{\zeta}),$$

which is (4.40). ∎

For the remainder of section 4, we bound $\|\Delta_{TS}\|_F$, $\|\delta X\|_F$, and $\|\Gamma_{TR}\|_F$ for MGS3 and BMGS_H and thus use the bounds (4.38)–(4.39) to establish the conditional backward error relationship (4.12).

4.2. Error analysis of MGS2. MGS2 is Schreiber and Van Loan's [32] algorithm for representing products of Householder transformations applied to the Q-R factorization of $\overline{X}$ in (1.2). From [32] and the analysis by Bischof and Van Loan [7], we may conclude that

$$\overline{X} + \Delta\overline{X} = U \begin{pmatrix} R \\ 0_{m\times n} \end{pmatrix}, \tag{4.41}$$

where

$$\|\Delta\overline{X}\|_F \le \varepsilon_M g_1(m,p)\|X\|_F + \mathcal{O}(\varepsilon_M^2), \tag{4.42}$$

for a modestly growing function $g_1(\cdot)$ ($g_1(\cdot)$ is not specified in [7, 32]). *The backward error* $\Delta\overline{X}$ *in* (4.41) *is distinct from* $\delta\overline{X}$ *in* (4.12) The authors of [7, 32] also show that U satisfies

$$\|I_m - U^T U\|_F \le \varepsilon_M g_2(m,n) + \mathcal{O}(\varepsilon_M^2) \tag{4.43}$$

for a modest sized function $g_2(\cdot)$ (also unspecified in [7, 32]).

Equation (4.42) is sufficient to show (4.5) and (4.6) as shown below. To show (4.6), we note that

$$\Gamma_{TR} = (I_n - T)R = \Delta\overline{X}(1\colon n, :\,),$$

and thus,

$$\|\Gamma_{TR}\|_F = \|\Delta\overline{X}(1\colon n, :\,)\|_F \le \varepsilon_M g_1(m,n)\|X\|_F + \mathcal{O}(\varepsilon_M^2). \tag{4.44}$$

Likewise, to show (4.5), we note that

$$\begin{aligned}
\delta X &= QR - X \\
&= Q(I_n - T)R + QTR - X \\
&= Q\Delta\overline{X}(1\colon n, :\,) + \Delta\overline{X}(n+1\colon m+n, :\,) \\
&= \begin{pmatrix} Q & I_m \end{pmatrix} \Delta\overline{X}.
\end{aligned}$$

Thus,

$$\begin{aligned}\|\delta X\|_F &\le \| \begin{pmatrix} Q & I_m \end{pmatrix} \|_2 \|\Delta \overline{X}\|_F \\ &= \left(1 + \|Q\|_2^2\right)^{1/2} \|\Delta \overline{X}\|_F \\ &\le \varepsilon_M \, (n+1)^{1/2} \, g_1(m,n) \|X\|_F. \end{aligned} \tag{4.45}$$

The bound (4.43) is sufficient to establish that MGS2 produces a stable factorization. However, because of the role MGS2 plays in MGS3, we need the following theorem.

THEOREM 4.7. *Let $T_k, k = 1, \ldots, n$, let R be computed as in MGS2, let S_k be given by* (2.16)*, and let R satisfy* (4.35)*. Then,*

$$T_k S_k = I_k + \Delta_k, \quad \|\Delta_k\|_F \le L_{TS}(m,k)\varepsilon_M + \mathcal{O}(\varepsilon_M^2), \tag{4.46}$$

where $L_{TS}(m,k) = \sqrt{1.5}mk$. Thus, for MGS2, Δ_{TS} in (4.1) *satisfies*

$$\|\Delta_{TS}\|_F \le \varepsilon_M L_{TS}(m,n) + \mathcal{O}(\varepsilon_M^2). \tag{4.47}$$

Proof. This is a simple induction argument. For $k = 1$, we note that

$$T_1 = S_1 = (1),$$

and thus (4.46) holds with $\Delta_1 = (0)$. For $k = 2, \ldots, s$, using error analysis bounds on basic operations in [23, pp. 67–73], we have that

$$\mathbf{g}_k = -T_{k-1}\widehat{Q}_{k-1}^T \mathbf{q}_k - T_{k-1}(\delta \mathbf{g}_k^{(1)}) - \delta \mathbf{g}_k^{(2)},$$

where

$$\begin{aligned}&\|\delta \mathbf{g}_k^{(1)}\|_2 \le m\varepsilon_M \|\widehat{Q}_{k-1}\|_F \|\mathbf{q}_k\|_2 + \mathcal{O}(\varepsilon_M^2) = m\sqrt{k-1}\varepsilon_M + \mathcal{O}(\varepsilon_M^2), \\ &\|\delta \mathbf{g}_k^{(2)}\|_2 \le k\varepsilon_M \|T_{k-1}\|_F \|\widehat{Q}_{k-1}^T \mathbf{q}_k\|_2 + \mathcal{O}(\varepsilon_M^2) \le 2(k-1)^{3/2}\varepsilon_M + \mathcal{O}(\varepsilon_M^2).\end{aligned}$$

Thus,

$$\mathbf{g}_k + \delta \mathbf{g}_k = -T_{k-1}\widehat{Q}_{k-1}^T \mathbf{q}_k,$$

where

$$\begin{aligned}\|\delta \mathbf{g}_k\|_2 &\le \|T_{k-1}\|_2 \|\delta \mathbf{g}_k^{(1)}\|_2 + \|\delta \mathbf{g}_k^{(2)}\|_2 + \mathcal{O}(\varepsilon_M^2) \\ &\le (m + 2(k-1))\sqrt{k-1}\varepsilon_M + \mathcal{O}(\varepsilon_M^2).\end{aligned}$$

Using the induction hypothesis,

$$\begin{aligned}T_k S_k &= \begin{pmatrix} T_{k-1} & \mathbf{g}_k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} S_{k-1} & \widehat{Q}_{k-1}^T \mathbf{q}_k \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} I_{k-1} + \Delta_k & T_{k-1}\widehat{Q}_{k-1}^T \mathbf{q}_k + \mathbf{g}_k \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} I_{k-1} + \Delta_{k-1} & \delta \mathbf{g}_k \\ 0 & 1 \end{pmatrix}.\end{aligned}$$

Therefore,

$$T_k S_k - I_k = \Delta_k = \begin{pmatrix} \Delta_{k-1} & \delta \mathbf{g}_k \\ 0 & 0 \end{pmatrix}$$

implying that

$$\|\Delta_k\|_F^2 = \|\Delta_{k-1}\|_F^2 + \|\delta \mathbf{g}_k\|_2^2.$$

That recursion yields

$$\begin{aligned} \|\Delta_k\|_F^2 &\le \left(\sum_{j=1}^{k-1} m^2(j-1) + 4(j-1)^3 \right) \varepsilon_M^2 \\ &= [m^2(k-2)(k-1)/2 + k^2(k-1)^2]\varepsilon_M^2 + \mathcal{O}(\varepsilon_M^3) \\ &< m^2 k^2 [1/2 + k^2/m^2]\varepsilon_M^2 + \mathcal{O}(\varepsilon_M^3) \le 1.5 m^2 k^2 \varepsilon_M^2 + \mathcal{O}(\varepsilon_M^3). \end{aligned}$$

Taking square roots yields the bound (4.46). Equation (4.47) is merely the case $k = n$. ∎

We note that MGS2 is just MGS3 for $p = 1$, and (4.44), (4.45), and (4.47) establish that bounds (4.4)–(4.6) hold with

$$f_{TS}(m,n,1) = L_{TS}(m,p) = \sqrt{1.5}mn, \quad f_X(m,n,1) = (n+1)^{1/2} g_1(m,n),$$

and

$$f_{TR}(m,n,1) = g_1(m,n).$$

4.3. Backward error bounds on MGS3 and BMGS_H. The first two theorems in this section prove the bounds (4.4)–(4.6) for MGS3 and BMGS_H. Our last theorem uses the perturbation theory in section 4.1 to establish that these algorithms produce Q and R that—for an exactly left orthogonal matrix V—satisfy bounds on $\|I_n - Q^T Q\|_F$ and $\|X - VR\|_F$ similar to those for MGS given in [8, 11].

The argument to produce these results works because steps (2) and (7) of MGS3 and BMGS_H produce Q_k, R_{kk}, and T_{kk} that satisfy bounds of the form producing Q_k, R_{kk}, and T_{kk} such that

$$T_{kk} S_{kk} - I_p = \Delta_{kk}, \quad \|\Delta_{kk}\|_F \le \varepsilon_M L_{TS}(m,p) + \mathcal{O}(\varepsilon_M^2), \tag{4.48}$$

$$Y_k + \Delta Y_k = Q_k R_{kk}, \quad \|\Delta Y_k\|_F \le \varepsilon_M L_Y(m,p)\|Y_k\|_F + \mathcal{O}(\varepsilon_M^2), \tag{4.49}$$

$$(I_p - T_{kk})R_{kk} = \Gamma_{kk}, \quad \|\Gamma_{kk}\|_F \le \varepsilon_M L_{TR}(m,p)\|Y_k\|_F + \mathcal{O}(\varepsilon_M^2), \tag{4.50}$$

where $L_{TS}(m,p)$ $L_Y(m,p)$, and $L_{TR}(m,p)$ are modestly growing functions and S_{kk} is the matrix defined by (3.9).

For MGS3, these steps are performed by MGS2. In section 4.2, we state the appropriate bounds based on analysis in [32, 7] and Theorem 4.7. From Theorem 4.7, equation (4.48) holds with $L_{TS}(m,p) = \sqrt{1.5}mp$; from (4.45), equation (4.49) holds with $L_Y(m,p) = (p+1)^{1/2} g_1(m,p)$; and from (4.44), equation (4.50) holds with $L_{TR}(m,p) = g_1(m,p)$.

For BMGS_H, these steps are performed with Householder Q-R factorization and $T_{kk} = I_p$. From [23, p. 360], we have that

$$\|I_p - Q_k^T Q_k\|_F \le d_Q m p^{3/2} \varepsilon_M + \mathcal{O}(\varepsilon_M^2),$$

and thus

$$\|T_{kk} S_{kk} - I_p\|_F = \|S_{kk} - I_p\|_F = \|\mathbf{triu}(Q_k^T Q_k - I_p)\|_F \le \varepsilon_M d_Q m p^{3/2} + \mathcal{O}(\varepsilon_M^2),$$

and (4.48) is satisfied with $L_{TS}(m,p) = d_Q mp^{3/2}$. From [23, Theorem 19.4], (4.49) holds with $L_Y(m,p) = d_Y mp$, where d_Y is a constant. Since $T_{kk} = I_p$, we have

$$\Gamma_{kk} = (I_p - T_{kk})R_{kk} = 0,$$

so that (4.50) holds with $L_{TR}(m,p) \equiv 0$.

This leads to Theorems 4.8 and 4.9. The hypotheses of these theorems specifically refer to "MGS3 or BMGS_H," *but their results hold if any factorization method that satisfies* (4.48)–(4.50) *is substituted for the Q-R factorization in steps* (2) *and* (7) *of MGS*3. Both theorems are proved in Appendix A.

The following integer functions will be used to establish the backward error bounds. In the definitions below, $g_1(m,p)$ is the function discussed after (4.42).

The first three functions are for (4.48)–(4.50):

$$L_{TS}(m,p) = \begin{cases} \sqrt{1.5}mp & \text{for MGS3,} \\ d_Q mp^{3/2} & \text{for BMGS\_H,} \end{cases} \tag{4.51}$$

$$L_Y(m,p) = \begin{cases} (p+1)^{1/2} g_1(m,p) & \text{for MGS3,} \\ d_Y mp & \text{for BMGS\_H,} \end{cases} \tag{4.52}$$

$$L_{TR}(m,p) = \begin{cases} g_1(m,p) & \text{for MGS3,} \\ 0 & \text{for BMGS\_H.} \end{cases} \tag{4.53}$$

The next three functions are for error bounds on matrix operations in the main loop of MGS3 and Function 3.3 (BMGS_H):

$$L_1(m,t) = (m+8t)t^{1/2}, \tag{4.54}$$

$$L_2(m,t) = \sqrt{2}mt^{1/2}, \tag{4.55}$$

$$L_3(m,t,p) = 8t^2 + 4mt^{1/2}p + 8p^2. \tag{4.56}$$

The function $f_{TS}(\cdot)$ is for bounding the backward error as in (4.4):

$$f_{TS}(m,n,p) = \sqrt{69s}\max\{L_3(m,n-p,p), L_{TS}(m,p)\}, \quad s = n/p. \tag{4.57}$$

The next three functions are for intermediate quantities in the error analysis proofs:

$$\begin{aligned} L_4(m,t,p) &= L_3(m,t,p) + 8L_{TR}(m,p) + 4L_Y(m,p) \\ &\quad + 4L_2(m,t) + 5f_{TS}(m,t,p) + 4L_1(m,t), \end{aligned} \tag{4.58}$$

$$L_{XR}(m,n,p) = 2\sqrt{10}f_{TS}(m,n,p) + \sqrt{5}L_1(m,t) + L_2(m,t), \tag{4.59}$$

$$L_{YR}(m,p) = L_Y(m,p) + (p+1)^{1/2}\left(L_{TR}(m,p) + 2L_{TS}(m,p)\right). \tag{4.60}$$

The last two functions are for bounding the backward errors in (4.5)–(4.6) and $s = n/p$:

$$f_X(m,n,p) = \sqrt{s}(L_Y(m,p) + L_2(m,p)), \tag{4.61}$$

$$f_{TR}(m,n,p) = \sqrt{2s}\max\{L_4(m,n-p,p), L_{TR}(m,p)\}, \tag{4.62}$$

$$f_{VR}(m,n,p) = \sqrt{2}f_X(m,n,p) + \sqrt{10}f_{TR}(m,n,p) + 2f_{TS}(m,n,p), \tag{4.63}$$

$$f_{UR}(m,n,p) = f_X(m,n,p) + \sqrt{5}(f_{TR}(m,n,p) + 2f_{TS}(m,n,p)). \tag{4.64}$$

The first theorem of this subsection establishes (4.4).

THEOREM 4.8. *Suppose that* $X \in \mathbf{R}^{m\times n}$ *is partitioned according to* (2.1) *and that its Q-R factorization is computed by MGS*3 *or BMGS_H. For* $k = 1, \ldots, s$, *let* T_k *be as produced by MGS*3 *or BMGS_H in floating point arithmetic with machine unit* ε_M. *Let* S_k *be given by* (2.16), *and assume that, at each step, the Q-R factorization of* Y_k *produces* T_{kk} *satisfying* (4.50). *Then,*

$$T_k S_k = I_{kp} + \Delta_k, \quad \|\Delta_k\|_2 \le \varepsilon_M f_{TS}(m, kp, p) + \mathcal{O}(\varepsilon_M^2), \tag{4.65}$$

where $f_{TS}(m, kp, p)$ *is given by* (4.57). *Thus, for* $k = s$ *we have the inequality* (4.4). *For the matrix* U *in* (2.17), *there exists an exactly orthogonal matrix* $\widetilde{U}$ *such that*

$$\|U - \widetilde{U}\|_F \le \varepsilon_M \sqrt{10n} f_{TS}(m, n, p) + \mathcal{O}(\varepsilon_M^2). \tag{4.66}$$

The second theorem of this subsection establishes (4.5)–(4.6).

THEOREM 4.9. *Assume the hypothesis and notation of Theorem* 4.8. *Assume that* R *is nonsingular and that* (4.35) *holds. For* $k = 1, \ldots, s$, *let* $\widehat{X}_k$ *and* $\widehat{Q}_k$ *be given by* (2.2), *and let* R_k *be given by* (3.3). *Then* $\widehat{Q}_k$, R_k, *and* T_k *satisfy*

$$\widehat{X}_k + \delta\widehat{X}_k = \widehat{Q}_k R_k, \tag{4.67}$$

$$\|\delta\widehat{X}_k\|_F \le \varepsilon_M f_X(m, kp, p)\|\widehat{X}_k\|_F + \mathcal{O}(\varepsilon_M^2), \tag{4.68}$$

$$(I_{n_k} - T_k)R_k = \Gamma_k, \quad \|\Gamma_k\|_F \le \varepsilon_M f_{TR}(m, kp, p)\|\widehat{X}_k\|_F + \mathcal{O}(\varepsilon_M^2). \tag{4.69}$$

Thus, from (4.68)–(4.69) *for* $k = s$, *we have* (4.5) *and* (4.6).

The final theorem establishes bounds for the loss of orthogonality in Q and loss of orthogonal similarity between X and R. These results establish that if a BMGS algorithm computes the Q-R factorization of X_1 and $Y_k, k = 2, \ldots, s$ with a TSQR factorization satisfying (4.48)–(4.50), then it has error analysis properties similar to those of MGS given in [8, 11].

THEOREM 4.10. *Assume the hypothesis and terminology of Theorem* 4.9. *Then* Q *satisfies*

$$\|I_n - Q^T Q\|_F \le 2\hat{\zeta}/(1 - \hat{\zeta}),$$
$$\hat{\zeta} = \varepsilon_M f_\zeta(m, n, p)\|X\|_F\|R^{-1}\|_2 + \mathcal{O}(\varepsilon_M^2),$$

where $f_\zeta(\cdot)$ *is given by* $\hat{\zeta}$ (4.34). *There exists a left orthogonal matrix* V *such that*

$$\|X - VR\|_F \le \varepsilon_M f_{VR}(m, n, p)\|X\|_F + \mathcal{O}(\varepsilon_M^2), \tag{4.70}$$

where $f_{VR}(\cdot)$ *is given by* (4.63). *Also, there exists an orthogonal* $\widetilde{U}$ *satisfying* (4.66) *and* (4.12) *with*

$$\|\delta\overline{X}\|_F \le \varepsilon_M f_{UR}(m, n, p)\|X\|_F + \mathcal{O}(\varepsilon_M^2), \tag{4.71}$$

where $f_{UR}(\cdot)$ *is given by* (4.64).

Proof. From (4.22),

$$X - VR = -FZ_1^T(\delta\overline{X}_1) - \delta\overline{X}_2$$

so that

$$\begin{aligned}\|X - VR\|_F &\le \|F\|_2\|Z_1\|_2\|\delta\overline{X}_1\|_F + \|\delta\overline{X}_2\|_F \\ &\le \|\delta\overline{X}_1\|_F + \|\delta\overline{X}_2\|_F \\ &\le \sqrt{2}\|\delta\overline{X}\|_F. \end{aligned} \tag{4.72}$$

Using (4.39) and Theorems 4.8 and 4.9, this becomes

$$\begin{aligned}\|X - VR\|_F &\le \sqrt{2}(\|\delta X\|_F + \sqrt{5}(\|\Gamma_{TR}\|_F + 2\|\Delta_{TS}\|_F))\|R\|_2 \\ &\le \varepsilon_M\sqrt{2}f_X(m,n,p)\|X\|_F + \varepsilon_M(\sqrt{10}f_{TR}(m,n,p)\|X\|_F \\ &\quad + 2\sqrt{2}f_{TS}(m,n,p)\|R\|_2) + \mathcal{O}(\varepsilon_M^2).\end{aligned}$$

Using (4.72) and orthogonal equivalence, we have that

$$\|R\|_2 \le \|X\|_F + \|X - VR\|_F$$

so that

$$\|X - VR\|_F \le \varepsilon_M f_{VR}(m,n,p)\|X\|_F + \varepsilon_M 2\sqrt{2}f_{TS}(m,n,p)\|X - VR\|_F + \mathcal{O}(\varepsilon_M^2).$$

Solving for $\|X - VR\|_F$ yields (4.70).

Equation (4.71) is just the result of (4.12) and (4.39) coupled with the bounds on $\|\Delta_{TS}\|_F$, $\|\Gamma_{TR}\|_F$, and $\|\delta X\|_F$ from Theorems 4.8 and 4.9. The bound on $\|I_n - Q^T Q\|_F$ results from Corollary 4.6. ∎

4.4. Numerical tests. We coded MGS, MGS2, MGS3, and BMGS_H in MATLAB 2013a on the author's Dell Inspiron 14z laptop. This is a sample of the tests that we did to assess the orthogonality of the computed Q and $Z = U(:, 1{:}\, n)$, where U is defined by (2.17). We let X be a 6000×1000 matrix and chose $p = 30$ for MGS3 and BMGS_H. The matrix X was constructed using the singular value decomposition into

$$X = U_X \Sigma_X V_X^T,$$

where $U_X \in \mathbf{R}^{6000\times 1000}$ is left orthogonal, $V_X \in \mathbf{R}^{1000\times 1000}$ is orthogonal, and $\Sigma_X \in \mathbf{R}^{1000\times 1000}$ is positive and diagonal. The matrices U_X and V_X were constructed using a method for constructing random orthogonal matrices described by Stewart [37]. The matrix Σ_X was given by

$$\Sigma_X = \mathrm{diag}(1, \delta, \delta^2, \ldots, \delta^{999}),$$

where $\delta^{999} = 1/\kappa$. The parameter κ had the values $\kappa = 10^t, t = 6, \ldots, 16$. Thus the condition number of X was set to

$$\kappa_2(X) = \|X^\dagger\|_2\|X\|_2 = \kappa.$$

For each of these values of κ, we computed

$$\mathrm{orth}_F(Q) = \|I_{1000} - Q^T Q\|_F$$

for all four algorithms. They are graphed on a $\log_{10}$-scale in Figure 1. The four algorithms (MGS, MGS2, MGS3, and BMGS_H) produced nearly identical orthogonality results; $\mathrm{orth}_F(Q)$ corresponds closely to $\varepsilon_M\kappa$ as would be expected.

We also computed $\mathrm{orth}_F(Z)$ for all four algorithms. Since all four algorithms produced $\mathrm{orth}_F(Z) \in (10^{-14}, 10^{-13})$ for all values of $\kappa_2(X)$, we graphed $\mathrm{orth}_F(Z)$ on a $\log_{10}$-scale for BMGS_H only and compared its orthogonality to that of Q produced by the Householder-based MATLAB **qr** function in Figure 2. As can be seen, the orthogonality of Z from BMGS_H and Q from MATLAB **qr** are indifferent to $\kappa_2(X)$ just as the results of this work predict; both are near left orthogonal.

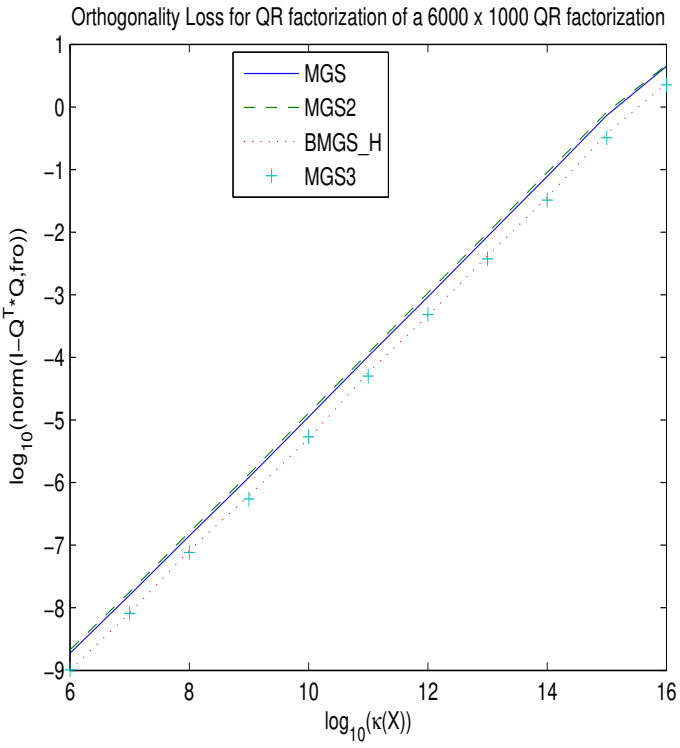

FIG. 1. *Orthogonality loss for Q relative to $\kappa_2(X)$.*

5. Previous BMGS formulations. Different formulations of block MGS are given by Jalby and Philippe [26] and Vanderstraeten [39]. We now relate them to MGS3 and BMGS_H. Function 5.1 (BMGS_JP) is the block MGS algorithm in [26].

FUNCTION 5.1 (Jalby–Philippe BMGS algorithm).
function $[Q, R] = \text{BMGS\_JP}(X, p)$
(1) $[m, n] = \mathbf{size}(X)$;$s = n/p$;
(2) $[Q_1, R_{11}] = \mathbf{MGS}(X_1)$
(3) **for** $k = 2\colon s$
(4) $\quad Y_k = X_k$;
(5) $\quad$ **for** $j = 1\colon k-1$
(6) $\quad\quad R_{jk} = Q_j^T Y_k$
(7) $\quad\quad Y_k = Y_k - Q_j R_{jk}$
(8) $\quad$ **end**;
(9) $\quad [Q_k, R_{kk}] = \mathbf{MGS}(Y_k)$;
(10) **end**;
(11) *Q is as in* (2.1) *and R is as in* (3.1)
(12) **end**. BMGS_JP

Although the Q-R factorization from this algorithm satisfies (4.2), the bound given by the authors of [26] for the orthogonality of Q is

$$\|I_n - Q^T Q\|_F \le \varepsilon_M f_{JP}(m,n)\kappa_2(R) \max_{1\le k\le s} \kappa_2(R_{kk})$$

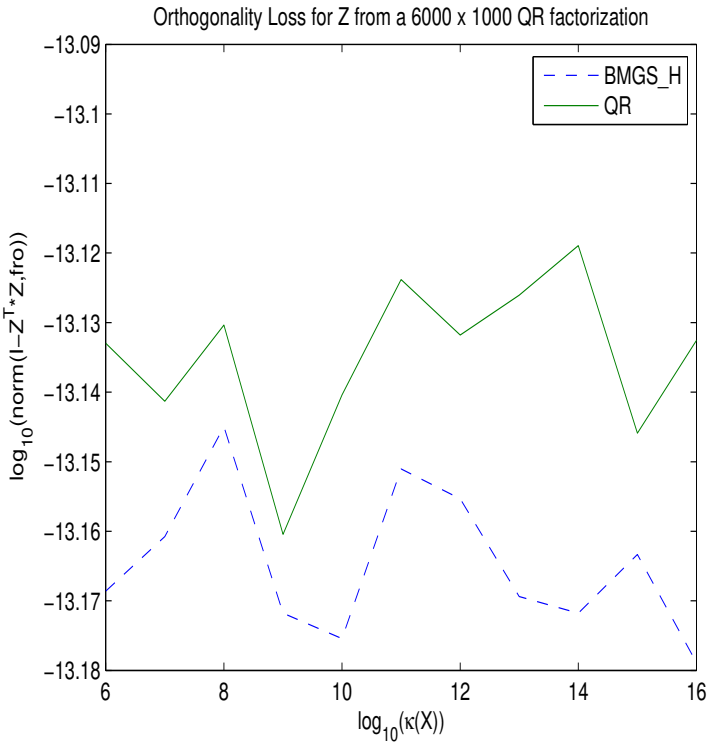

FIG. 2. *Orthogonality loss for* $Z = U(:, 1{:}\, n)$ *relative to* $\kappa_2(X)$.

for a modestly sized function $f_{JP}(m, n)$, where $\kappa_2(C) = \|C^{-1}\|_2\|C\|_2$ for a nonsingular matrix C. There is no guarantee that BMGS_JP produces a small value of $\|\Gamma_{TR}\|_F$ for Γ_{TR} in (4.3), so BMGS_JP does not fall under the framework described in section 4. However, that can be corrected by changing the manner in which Y_k is factored.

In contrast to MGS3 and BMGS_H, BMGS_JP takes the transition from MGS to MGS2 backward—it substitutes an inner loop for computing Y_k for the computation of T.

To produce a more nearly orthogonal Q, Vanderstraeten [39] replaces MGS in steps (2) and (9) with

$[\widetilde{Y}_k, R_{kk}^{(1)}] = \mathbf{MGS}(Y_k);$
$[Q_k, R_{kk}^{(2)}] = \mathbf{MGS}(\widetilde{Y}_k);$
$R_{kk} = R_{kk}^{(2)} R_{kk}^{(1)};$

Under a condition similar to (4.32), the resulting Q_k would be near left orthogonal. However, this is not the most efficient method for producing the TSQR needed here. Instead, we propose to replace statements (2) and (9) with

$$(2')\ [Q_1, R_{11}] = \mathbf{House\_QR}(X_1), \tag{5.1}$$

$$(9')\ [Q_k, R_{kk}] = \mathbf{House\_QR}(Y_k), \tag{5.2}$$

where **House_QR** is as in (3.13) and thus denotes a Householder-based TSQR

factorization. As in the standard version of BMGS_JP, we do not need to store the triangular matrix T, and, as in BMGS_H, we let $T_{kk} = I_p$.

In exact arithmetic, the relationship between BMGS_JP with the modifications (5.1)–(5.2) and BMGS_H follows from a simple argument. Let

$$H_k^{(j)} = \begin{pmatrix} R_{1k} \\ \vdots \\ R_{jk} \end{pmatrix}.$$

The first step is

$$H_k^{(1)} = R_{1k} = Q_1^T X_k = T_1^T \widehat{Q}_1^T X_k.$$

If we assume that

$$H_k^{(j-1)} = T_{j-1}^T \widehat{Q}_{j-1}^T X_k,$$

then

$$H_k^{(j)} = \begin{pmatrix} H_k^{(j-1)} \\ R_{jk} \end{pmatrix},$$

where

$$R_{jk} = Q_j^T Y_k^{(j-1)}, \quad Y_k^{(j-1)} = X_k - \widehat{Q}_{j-1} H_k^{(j-1)}.$$

Thus,

$$\begin{aligned} R_{jk} &= T_{jj}^T Q_j^T X_k - Q_j^T \widehat{Q}_{j-1} H_k^{(j-1)} \\ &= Q_j^T X_k - Q_j^T \widehat{Q}_{j-1} T_{j-1}^T \widehat{Q}_{j-1}^T X_k \\ &= \begin{pmatrix} -Q_j^T \widehat{Q}_{j-1} T_{j-1}^T & I_p \end{pmatrix} \begin{pmatrix} \widehat{Q}_{j-1} \\ Q_j \end{pmatrix}^T X_k \\ &= \begin{pmatrix} -Q_j^T \widehat{Q}_{j-1} T_{j-1}^T & I_p \end{pmatrix} \widehat{Q}_j^T X_k. \end{aligned}$$

The above may be summarized as

$$\begin{aligned} H_k^{(j)} &= \begin{pmatrix} T_{j-1}^T & 0 \\ -Q_j^T \widehat{Q}_{j-1} T_{j-1}^T & I_p \end{pmatrix} \widehat{Q}_j^T X_k \\ &= T_j^T \widehat{Q}_j^T X_k. \end{aligned}$$

If we let $j = k - 1$, we have

$$H_k = T_{k-1}^T \widehat{Q}_k^T X_k,$$

which is lines (4)–(5) of BMGS_H. Hence, the modifications suggested in (5.1)–(5.2) transform BMGS_JP into another implementation of BMGS_H. From the above derivation, we strongly conjecture that Jalby and Philippe's algorithm with the modifications (5.1)–(5.2) produces error analysis results similar to those in Theorems 4.8 and 4.9.

By an argument similar to that above, if we replace lines (2) and (9) of BMGS_JP with

$$(2')\ [Q_1, R_{11}, T_{11}] = \mathbf{MGS2}(X_1),$$
$$(9')\ [Q_k, R_{kk}, T_{kk}] = \mathbf{MGS2}(Y_k),$$

and replace (6) with

$$(6')\ R_{jk} = T_{jj}^T Q_j^T Y_k,$$

then BMGS_JP becomes another implementation of MGS3.

6. Conclusion. The MGS algorithm produces good least squares solutions because it has small bounds on the residual (4.2) and on the residuals (4.1) and (4.3) for the implied Householder Q-R factorization in section 2.3.

BMGS algorithms inherit the favorable error analysis properties of the MGS algorithm provided that they produce Q, R, and T such that all three residuals (4.1)–(4.3) satisfy bounds of the form (4.4)–(4.6). Such bounds ensure that the Sheffield structure from [28] is satisfied with a backward error of $\mathcal{O}(\varepsilon_M\|X\|_F)$ as shown in Theorem 4.1.

The BMGS algorithms in MGS3 and BMGS_H are shown to satisfy the bounds (4.4)–(4.6) because of how Y_k is computed from X_k in lines (4)–(6) of these procedures and because the Q-R factorizations of X_1 in line (2) and Y_k in line (7) are done by a procedure which satisfies the bounds (4.48)–(4.50). The final procedure, BMGS_H, is based entirely upon matrix-matrix operations and "tall, skinny" Q-R factorization. The structure described was shown in section 5 to be applicable to variants of the algorithm of Jalby and Philippe [26] and leads to a more inexpensive way to obtain a conditionally backward stable factorization.

Appendix A. Proof of Theorems 4.8 and 4.9. To set up the detailed, technical proofs of Theorems 4.8 and 4.9, we begin with a lemma (Lemma A.1) in section A.1 that bounds the error in the computed quantities H_k, Y_k, and G_k from one BMGS step in MGS3 and BMGS_H. In section A.2, the proof of Theorem 4.8 bounding $\|\Delta_{TS}\|_F$ in (4.1) follows directly from an induction argument on one of the bounds in Lemma A.1. In section A.3, the proof of Theorem 4.9 bounding $\|\delta X\|_F$ and $\|\Gamma_{TR}\|_F$ in (4.2)–(4.3) also needs the bounds in Lemma A.1. That proof also requires two technical lemmas: the first (Lemma A.2) bounds important quantities in terms of $\|X_k\|_F$; the second (Lemma A.3) is a key technical result necessary to bound $\|\Gamma_{TR}\|_F$. The proof of Theorem 4.9 is an induction argument from those two lemmas.

A.1. Error bounds for matrix operations. The first lemma needed to prove Theorems 4.8 and 4.9 uses the two simple floating bound error bounds for matrices $A, C \in \mathbf{R}^{m\times n}$ and $B \in \mathbf{R}^{n\times p}$. These are

$$\text{fl}(AB) = AB + E, \quad \|E\|_F \le n\|A\|_F\|B\|_F\varepsilon_M + \mathcal{O}(\varepsilon_M^2), \tag{A.1}$$

$$\text{fl}(A + C) = A + C + E, \quad \|E\|_F \le \|A + C\|_F\varepsilon_M. \tag{A.2}$$

Equation (A.1) is a version of the bound in [23, p. 71], and (A.2) is just the error in one floating point addition applied to all of the entries of A and C. The matrix operations whose error bounds are given in Lemma A.1 are from the main loop of MGS3 and BMGS_H.

LEMMA A.1. *Assume the hypothesis of Theorem* 4.9. *Also let* $t = (k-1)p$ *be the number of columns in* $\widehat{Q}_{k-1}$. *The intermediate quantities computed in the main loop of MGS*3 *and BMGS_H satisfy the following error bounds:*

$$H_k + \delta H_k = T_{k-1}^T\widehat{Q}_{k-1}^T X_k, \quad \|\delta H_k\|_F \le \varepsilon_M L_1(m,t)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2), \tag{A.3}$$

$$Y_k + \delta Y_k = X_k - \widehat{Q}_{k-1}H_k, \quad \|\delta Y_k\|_F \le \varepsilon_M L_2(m,t)\left\|\begin{pmatrix} H_k \\ Y_k \end{pmatrix}\right\|_F + \mathcal{O}(\varepsilon_M^2), \tag{A.4}$$

$$G_k + \delta G_k = -T_{k-1}\widehat{Q}_{k-1}^T Q_k T_{kk}, \quad \|\delta G_k\|_F \le \varepsilon_M L_3(m,t,p) + \mathcal{O}(\varepsilon_M^2), \tag{A.5}$$

where $L_1(m,t)$, $L_2(m,t)$, *and* $L_3(m,t,p)$ *are given in* (4.54)–(4.56).

Proof. The operations from (A.3)–(A.5) are combinations of (A.1) and (A.2). To prove (A.3), we have that

$$H_k = T_{k-1}^T \widehat{Q}_{k-1}^T X_k + T_{k-1}^T(\delta H_k^{(1)}) + \delta H_k^{(2)},$$

where $\delta H_k^{(1)}$ is the error from the operation $\widehat{Q}_{k-1}^T X_k$, and $\delta H_k^{(2)}$ is the error from multiplying T_{k-1}^T by the computed product of $\widehat{Q}_{k-1}^T X_k$. Thus,

$$\begin{aligned}
\|\delta H_k^{(1)}\|_F &\le m\varepsilon_M \|\widehat{Q}_{k-1}\|_F \|X_k\|_F + \mathcal{O}(\varepsilon_M^2) \le mt^{1/2}\varepsilon_M \|X_k\|_F + \mathcal{O}(\varepsilon_M^2),\\
\|\delta H_k^{(2)}\|_2 &\le t\varepsilon_M \|T_{k-1}\|_F \|\widehat{Q}_{k-1}^T X_k\|_F + \mathcal{O}(\varepsilon_M^2)\\
&\le t^{3/2}\varepsilon_M \|T_{k-1}\|_2 \|\widehat{Q}_{k-1}\|_2 \|X_k\|_F + \mathcal{O}(\varepsilon_M^2)\\
&\le 4t^{3/2}\varepsilon_M \|X_k\|_F + \mathcal{O}(\varepsilon_M^2).
\end{aligned}$$

Thus,

$$\begin{aligned}
\|\delta H_k\|_F &= \|T_{k-1}^T(\delta H_k^{(1)}) + \delta H_k^{(2)}\|_F\\
&\le \|T_{k-1}\|_2 \|\delta H_k^{(1)}\|_F + \|\delta H_k^{(2)}\|_F\\
&\le (8t+m)t^{1/2}\|X_k\|_F \varepsilon_M + \mathcal{O}(\varepsilon_M^2).
\end{aligned}$$

To prove (A.4), we have that the matrix multiply and add satisfies

$$Y_k = X_k - \widehat{Q}_{k-1} H_k + \delta Y_k^{(1)} + \delta Y_k^{(2)},$$

where

$$\|\delta Y_k^{(1)}\|_F \le m\varepsilon_M \|\widehat{Q}_{k-1}\|_F \|H_k\|_F + \mathcal{O}(\varepsilon_M^2) \le mt^{1/2}\|H_k\|_F + \mathcal{O}(\varepsilon_M^2)$$

and

$$\|\delta Y_k^{(2)}\|_F \le \varepsilon_M \|Y_k\|_F.$$

Thus,

$$\begin{aligned}
\|\delta Y_k\|_F &\le (mt^{1/2}\|H_k\|_F + \|Y_k\|_F)\varepsilon_M \ + \mathcal{O}(\varepsilon_M^2)\\
&\le \left(m^2 t + 1\right)^{1/2} \left\| \begin{pmatrix} \|H_k\|_F \\ \|Y_k\|_F \end{pmatrix} \right\|_F \varepsilon_M + \mathcal{O}(\varepsilon_M^2) \le \sqrt{2} mt^{1/2} \left\| \begin{pmatrix} H_k \\ Y_k \end{pmatrix} \right\|_F + \mathcal{O}(\varepsilon_M^2),
\end{aligned}$$

which is (A.4).

To show (A.5), we note that the three matrix multiplications are

$$G_k = -T_{k-1}\widehat{Q}_{k-1}^T Q_k T_{kk} + \delta G_k^{(1)} - T_{k-1}(\delta G_k^{(2)}) - T_{k-1}\widehat{Q}_{k-1}^T(\delta G_k^{(3)}),$$

where $\delta G_k^{(j)}$ are the errors from the three matrix multiplications involved in computing G_k. Thus, since $\|T_{kk}\|_F \le \sqrt{p}\|T_{kk}\|_2 \le \sqrt{p}\|T\|_2 \le 2\sqrt{p}$, we have

$$\begin{aligned}
\|\delta G_k^{(1)}\|_F &\le t\varepsilon_M \|T_{k-1}\|_F \|\widehat{Q}_{k-1}^T Q_k T_{kk}\|_F + \mathcal{O}(\varepsilon_M^2)\\
&\le t\varepsilon_M \|T_{k-1}\|_F \|T_{kk}\|_2 \|\widehat{Q}_{k-1}\|_F \|Q_k\|_2 + \mathcal{O}(\varepsilon_M^2)\\
&\le 8t^2\varepsilon_M + \mathcal{O}(\varepsilon_M^2),\\
\|\delta G_k^{(2)}\|_F &\le m\varepsilon_M \|\widehat{Q}_{k-1}\|_F \|Q_k T_{kk}\|_F + \mathcal{O}(\varepsilon_M^2)\\
&\le m\varepsilon_M \|\widehat{Q}_{k-1}\|_F \|Q_k\|_F \|T_{kk}\|_2 + \mathcal{O}(\varepsilon_M^2)\\
&\le 2mt^{1/2}p^{1/2}\varepsilon_M + \mathcal{O}(\varepsilon_M^2),\\
\|\delta G_k^{(3)}\|_2 &\le p\varepsilon_M \|Q_k\|_F \|T_{kk}\|_F \le 2p^2\varepsilon_M + \mathcal{O}(\varepsilon_M^2).
\end{aligned}$$

Thus we have (A.5) with $\|\delta G_k\|_2$ bounded by

$$\begin{aligned}\|\delta G_k\|_2 &\le \|\delta G_k^{(1)}\|_2 + \|T_{k-1}\|_2\|\delta G_k^{(2)}\|_2 + \|T_{k-1}\|_2\|\widehat{Q}_{k-1}\|_2\|\delta G_k^{(3)}\|_F \\ &\le (8t^2 + 4mt^{1/2}p^{1/2} + 8p^2)\varepsilon_M + \mathcal{O}(\varepsilon_M^2) = L_3(m,t,p)\varepsilon_M + \mathcal{O}(\varepsilon_M^2). \qquad \square\end{aligned}$$

A.2. Proof of Theorem 4.8. The bound on $\|\delta G_k\|_F$ in Lemma A.1 allows us to prove Theorem 4.8.

Proof of Theorem 4.8. This induction proof makes use of the definition of Δ_{kk} in (4.48). For $k = 1$, we have that

$$\begin{aligned}\Delta_1 &= T_1 S_1 - I_p \\ &= T_{11}S_{11} - I_p = \Delta_{11}.\end{aligned}$$

Thus,

$$\|\Delta_1\|_F = \|\Delta_{11}\|_F \le \varepsilon_M L_{TS}(m,p) + \mathcal{O}(\varepsilon_M^2).$$

For $k > 1$, let $t = (k-1)p$. We have

$$\begin{aligned}T_k S_k &= \begin{pmatrix} T_k & G_k \\ 0 & T_{kk}\end{pmatrix}\begin{pmatrix} S_{k-1} & \widehat{Q}_{k-1}^T Q_k \\ 0 & S_{kk}\end{pmatrix} \\ &= \begin{pmatrix} T_{k-1}S_{k-1} & T_{k-1}\widehat{Q}_{k-1}^T Q_k + G_k S_{kk} \\ 0 & T_{kk}S_{kk}\end{pmatrix} \\ &= \begin{pmatrix} I_t + \Delta_{k-1} & T_{k-1}\widehat{Q}_{k-1}^T Q_k - T_{k-1}\widehat{Q}_{k-1}^T Q_k T_{kk} S_{kk} - (\delta G_k)S_{kk} \\ 0 & I_p + \Delta_{kk}\end{pmatrix} \\ &= \begin{pmatrix} I_t + \Delta_{k-1} & -T_{k-1}\widehat{Q}_{k-1}^T Q_k \Delta_{kk} - (\delta G_k)S_{kk} \\ 0 & I_p + \Delta_{kk}\end{pmatrix}.\end{aligned}$$

Thus,

$$\begin{aligned}\Delta_k &= T_k S_k - I_{kp} \\ &= \begin{pmatrix} \Delta_{k-1} & -T_{k-1}\widehat{Q}_{k-1}^T Q_k \Delta_{kk} - (\delta G_k)S_{kk} \\ 0 & \Delta_{kk}\end{pmatrix}\end{aligned}$$

so that for $t = (k-1)p$,

$$\begin{aligned}\|\Delta_k\|_F^2 &\le \|\Delta_{k-1}\|_F^2 + \|\delta G_k\|_F^2\|S_{kk}\|_2^2 + \|\Delta_{kk}\|_F^2(1 + \|T_{k-1}\|_2^2\|\widehat{Q}_{k-1}\|_2^2\|Q_k\|_2^2) \\ &\le \varepsilon_M^2[f_{TS}(m,t,p)^2 + 4L_3^2(m,t,p) + 65L_{TS}^2(m,p)] + \mathcal{O}(\varepsilon_M^3).\end{aligned}$$

This recurrence is bounded by

$$\|\Delta_k\|_F^2 \le 69k\varepsilon_M^2 \max\{L_3^2(m,t,p), L_{TS}^2(m,p)\} + \mathcal{O}(\varepsilon_M^3).$$

Taking square roots yields (4.65). If we let $k = s = n/p$, then we have (4.4). If we just apply (4.11) to (4.4), we have (4.66). $\square$

A.3. Proof of Theorem 4.9. We now give necessary bounds on $\|\binom{H_k}{Y_k}\|_F$ in (2.1) and $\|R_{kk}\|_F$ from (3.3). These bounds simplify the proof of Theorem 4.9 by allowing us to state all of our error bounds in terms of $\|X_k\|_F$.

LEMMA A.2. *Assume the hypothesis and terminology of Lemma* A.1. *Let* X_k *and* R_k *be given by the partition* (2.1). *Then,*

$$\left\| \begin{pmatrix} H_k \\ Y_k \end{pmatrix} \right\|_F \leq (1 + \varepsilon_M L_{XR}(m,t,p))\|X_k\|_F + \mathcal{O}(\varepsilon_M^2), \quad t = (k-1)p, \tag{A.6}$$

where $L_{XR}(m,t,p)$ *is given by* (4.59). *Thus,*

$$\|\delta Y_k\|_F \leq L_2(m,t)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2). \tag{A.7}$$

If we assume that the Q-R factorization of Y_k *satisfies* (4.48)–(4.50), *then*

$$\|R_{kk}\|_F \leq (1 + L_{YR}(m,p)\varepsilon_M)\|Y_k\|_F + \mathcal{O}(\varepsilon_M^2), \tag{A.8}$$

where $L_{YR}(m,p)$ *is defined by* (4.60).

Proof. Using the definition of U_{k-1}, the results of Lemma A.1 can be written

$$U_{k-1}^T \begin{pmatrix} 0 \\ X_k \end{pmatrix} = \begin{pmatrix} H_k \\ Y_k \end{pmatrix} + \begin{pmatrix} \delta H_k \\ \delta Y_k + \widehat{Q}_{k-1}\delta H_k \end{pmatrix}. \tag{A.9}$$

By the induction hypothesis and the use of Theorem 4.1, there is an exactly orthogonal matrix $\widetilde{U}_{k-1}$ such that

$$\|U_{k-1} - \widetilde{U}_{k-1}\|_F \leq \sqrt{10}\|Q_{k-1}\|_2\|\Delta_{k-1}\|_F \leq 2\sqrt{10} f_{TS}(m,t,p)\varepsilon_M + \mathcal{O}(\varepsilon_M^2).$$

Thus, we may write (A.9) as

$$\begin{pmatrix} H_k \\ Y_k \end{pmatrix} = \widetilde{U}_{k-1}^T \begin{pmatrix} 0 \\ X_k \end{pmatrix} - \begin{pmatrix} \delta H_k \\ \delta Y_k - \widehat{Q}_{k-1}\delta H_k \end{pmatrix} + (U_{k-1} - \widetilde{U}_{k-1})^T \begin{pmatrix} 0 \\ X_k \end{pmatrix}.$$

Bounding the norm of the left side of this equation with the right, we have

$$\begin{aligned}
\left\| \begin{pmatrix} H_k \\ Y_k \end{pmatrix} \right\|_F &\leq \|X_k\|_F + \|\delta Y_k\|_F + \left\| \begin{pmatrix} I_t \\ \widehat{Q}_{k-1} \end{pmatrix} \right\|_2 \|\delta H_k\|_F + \|U_{k-1} - \widetilde{U}_{k-1}\|_F \|X_k\|_F \\
&\leq \|X_k\|_F + \varepsilon_M L_2(m,t) \left\| \begin{pmatrix} H_k \\ Y_k \end{pmatrix} \right\|_F + \varepsilon_M \left(1 + \|\widehat{Q}_{k-1}\|_2^2\right)^{1/2} L_1(m,t)\|X_k\|_F \\
&\quad + \|\widehat{Q}_{k-1}\|_2 \sqrt{10} f_{TS}(m,t,p)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2) \\
&\leq \|X_k\|_F + \varepsilon_M L_2(m,t) \left\| \begin{pmatrix} H_k \\ Y_k \end{pmatrix} \right\|_F + \varepsilon_M \sqrt{5} L_1(m,t)\|X_k\|_F \\
&\quad + 2\sqrt{10}\varepsilon_M f_{TS}(m,t,p)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2).
\end{aligned}$$

Solving for $\left\| \begin{pmatrix} H_k \\ Y_k \end{pmatrix} \right\|_F$ yields

$$\begin{aligned}
\left\| \begin{pmatrix} H_k \\ Y_k \end{pmatrix} \right\|_F &\leq \|X_k\|_F (1 + \varepsilon_M \sqrt{5} L_1(m,t) \\
&\qquad + 2\varepsilon_M \sqrt{10} f_{TS}(m,t,p))/(1 - \varepsilon_M L_2(m,t)) + \mathcal{O}(\varepsilon_M^2) \\
&\leq (1 + \varepsilon_M(\sqrt{5} L_1(m,t) + L_2(m,t) + 2\sqrt{10} f_{TS}(m,t,p)))\|X_k\|_F + \mathcal{O}(\varepsilon_M^2) \\
&= (1 + \varepsilon_M L_{XR}(m,t,p))\|X_k\|_F + \mathcal{O}(\varepsilon_M^2).
\end{aligned}$$

Equation (A.7) is just (A.4) combined with (A.6).

If we invoke Theorem 4.1 and use (4.48)–(4.50), there is an exactly orthogonal matrix $\widetilde{U}_{kk}$ such that

$$\begin{pmatrix} 0_{p\times p} \\ Y_k \end{pmatrix} + \delta\overline{Y}_k = \widetilde{U}_{kk} \begin{pmatrix} R_{kk} \\ 0_{m\times p} \end{pmatrix}, \tag{A.10}$$

where

$$\begin{aligned} \|\delta\overline{Y}_k\|_F &\le \|\Delta Y_k\|_F + \left(1 + \|Q_k\|_2^2\right)^{1/2} \left(\|\Gamma_{kk}\|_F + 2\|\Delta_{kk}\|_F \|R_{kk}\|_2\right) \\ &\le \|\Delta Y_k\|_F + (p+1)^{1/2} \left(\|\Gamma_{kk}\|_F + 2\|\Delta_{kk}\|_F \|R_{kk}\|_F\right) \\ &\le \varepsilon_M (L_Y(m,p) + (p+1)^{1/2} L_{TR}(m,p)) \|Y_k\|_F \\ &\quad + 2\varepsilon_M (p+1)^{1/2} L_{TS}(m,p) \|R_{kk}\|_F + \mathcal{O}(\varepsilon_M^2). \end{aligned} \tag{A.11}$$

Thus, using (A.11) to bound $\|R_{kk}\|_F$ we have

$$\begin{aligned} \|R_{kk}\|_F &\le \|Y_k\|_F + \|\delta\overline{Y}_k\|_F \\ &\le (1 + \varepsilon_M((L_Y(m,p) + (p+1)^{1/2} L_{TR}(m,p)) \|Y_k\|_F) \\ &\quad + 2\varepsilon_M (p+1)^{1/2} L_{TS}(m,p)) \|R_{kk}\|_F + \mathcal{O}(\varepsilon_M^2). \end{aligned} \tag{A.12}$$

Solving for $\|R_{kk}\|_F$ yields

$$\begin{aligned} \|R_{kk}\|_F &\le \|Y_k\|_F (1 + \varepsilon_M((L_Y(m,p) \\ &\quad + (p+1)^{1/2} L_{TR}(m,p)) \|Y_k\|_F) / (1 - 2\varepsilon_M (p+1)^{1/2} L_{TS}(m,p))) + \mathcal{O}(\varepsilon_M^2) \\ &= (1 + \varepsilon_M (L_Y(m,p) + (p+1)^{1/2} (L_{TR}(m,p) + 2L_{TS}(m,p))) \|Y_k\|_F + \mathcal{O}(\varepsilon_M^2) \\ &= (1 + \varepsilon_M L_{YR}(m,p)) \|Y_k\|_F + \mathcal{O}(\varepsilon_M^2). \end{aligned}$$

∎

Before proving Theorem 4.9, we prove a key technical lemma that links the computation of T with the computation of R in MGS3 and BMGS_H.

LEMMA A.3. *Assume the hypothesis and notation of Lemma* A.1. *Assume that* $T_{kk}, R_{kk} \in \mathbf{R}^{p\times p}$ *satisfy* (4.50) *and that* Q_k *and* R_{kk} *satisfy* (4.49). *Then, letting* $t = (k-1)p$, *we have*

$$\|(I - T_{k-1})H_k - G_k R_{kk}\|_F \le \varepsilon_M L_4(m,t,p) \|X_k\|_F + \mathcal{O}(\varepsilon_M^2), \tag{A.13}$$

where $L_4(\cdot)$ *is given by* (4.58).

Proof. This proof is a combination of our assumptions about the Q-R factorization in steps (2) and (7) of MGS3 and BMGS_H, the error bounds in Lemma A.1, and Theorem 4.8.

From (A.5), we have that

$$\begin{aligned} G_k R_{kk} &= -T_{k-1} \widehat{Q}_{k-1}^T Q_k T_{kk} R_{kk} + E_1, \\ E_1 &= -(\delta G_k) R_{kk}. \end{aligned}$$

We then use our assumptions about T_{kk} and R_{kk} to show that

$$\begin{aligned} G_k R_{kk} &= -T_{k-1} \widehat{Q}_{k-1}^T Q_k R_{kk} + T_{k-1} \widehat{Q}_{k-1}^T Q_k (I_p - T_{kk}) R_{kk} + E_1 \\ &= -T_{k-1} \widehat{Q}_{k-1}^T Q_k R_{kk} + E_2 + E_1, \\ E_2 &= T_{k-1} \widehat{Q}_{k-1}^T Q_k \Gamma_{kk}. \end{aligned}$$

Using our assumption about the backward error in the Q-R factorization of Y_k and the error (A.4), we have

$$G_k R_{kk} = -T_{k-1}\widehat{Q}_{k-1}^T(X_k - \widehat{Q}_{k-1}H_k) + E_3 + E_2 + E_1,$$
$$E_3 = T_{k-1}\widehat{Q}_{k-1}(\delta Y_k - \Delta Y_k).$$

The definition of S_{k-1} in (2.16) expands this into

$$\begin{aligned} G_k R_{kk} &= -T_{k-1}\widehat{Q}_{k-1}^T X_k + T_{k-1}(S_{k-1} + S_{k-1}^T - I)H_k + E_3 + E_2 + E_1 \\ &= -T_{k-1}S_{k-1}^T T_{k-1}^T \widehat{Q}_{k-1}^T X_k + T_{k-1}(S_{k-1} + S_{k-1}^T - I)H_k + E_4 + E_3 + E_2 + E_1, \\ E_4 &= T_{k-1}\Delta_{k-1}^T \widehat{Q}_{k-1}^T X_k. \end{aligned}$$

Finally, we make use of the bound (A.3) to reveal the desired relationship between the computation of G_k to form T_k and the computation of H_k to form R_k. We have

$$\begin{aligned} G_k R_{kk} &= -T_{k-1}S_{k-1}^T(H_k + \delta H_k) + T_{k-1}(S_{k-1} + S_{k-1}^T - I)H_k \\ &\quad + E_4 + E_3 + E_2 + E_1 \\ &= (I_t - T_{k-1})H_k + E_6 + E_5 + E_4 + E_3 + E_2 + E_1 \qquad \text{(A.14)} \\ E_5 &= -T_{k-1}S_{k-1}^T(\delta H_k), \quad E_6 = (T_{k-1}S_{k-1} - I)H_k = \Delta_{k-1}H_k. \end{aligned}$$

Thus,

$$G_k R_{kk} - (I_t - T_{k-1})H_k = E,$$

where

$$E = E_6 + E_5 + E_4 + E_3 + E_2 + E_1.$$

Our bounds on the six error matrices are

$$\begin{aligned} \|E_1\|_F &\le \|\delta G_k\|_F \|R_{kk}\|_F \le \varepsilon_M L_3(m,t,p)\|R_{kk}\|_F + \mathcal{O}(\varepsilon_M^2) \\ &\le \varepsilon_M L_3(m,t,p)\|Y_k\|_F + \mathcal{O}(\varepsilon_M^2) \le \varepsilon_M L_3(m,t,p)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2), \\ \|E_2\|_F &\le \|T_{k-1}\|_2 \|\widehat{Q}_{k-1}\|_2 \|Q_k\|_2 \|\Gamma_{kk}\|_F \le 8\|\Gamma_{kk}\|_F \\ &\le 8\varepsilon_M L_{TR}(m,p)\|Y_k\|_F + \mathcal{O}(\varepsilon_M^2) \le 8\varepsilon_M L_{TR}(m,p)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2), \\ \|E_3\|_F &\le \|T_{k-1}\|_2 \|\widehat{Q}_{k-1}\|_2 (\|\delta Y_k\|_F + \|\Delta Y_k\|_F) \\ &\le 4\varepsilon_M [L_Y(m,p)\|Y_k\|_F + L_2(m,t)\|X_k\|_F] + \mathcal{O}(\varepsilon_M^2) \\ &\le 4\varepsilon_M [L_Y(m,p) + L_2(m,t)]\|X_k\|_F + \mathcal{O}(\varepsilon_M^2), \\ \|E_4\|_F &\le \|T_{k-1}\|_2 \|\widehat{Q}_{k-1}\|_2 \|\Delta_{k-1}\|_2 \|X_k\|_F \le 4\varepsilon_M f_{TS}(m,t,p)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2), \\ \|E_5\|_F &\le \|T_{k-1}\|_2 \|S_{k-1}\|_2 \|\delta H_k\|_F \le 4\varepsilon_M L_1(m,t)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2), \\ \|E_6\|_F &\le \|H_k\|_2 \|\Delta_{k-1}\|_F \le \varepsilon_M f_{TS}(m,t,p)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2). \end{aligned}$$

Thus,

$$\begin{aligned} \|E\|_2 &\le \|E_1\|_F + \|E_2\|_F + \|E_3\|_F + \|E_4\|_F + \|E_5\|_F + \|E_6\|_F \\ &= \varepsilon_M L_4(m,t,p)\|X_k\|_F + \mathcal{O}(\varepsilon_M^2), \end{aligned}$$

where $L_4(m,t,p)$ is given by (4.58). ∎

The cancellation of the term $T_{k-1}S_{k-1}^T H_k$ in (A.14), necessary to complete the proof of Lemma A.3, occurs because Y_k is computed according to (3.7). It is the only argument in the proof of either Theorem 4.8 or Theorem 4.9 that cannot be made for classical Gram–Schmidt.

Proof of Theorem 4.9. First, we prove the simpler bound, (4.67)–(4.68), by a simple induction argument. For $k = 1$, we have

$$\widehat{Q}_1 R_1 = Q_1 R_1 = X_1 + \delta X_1, \quad \|\delta X_1\|_F \leq \varepsilon_M L_Y(m,p)\|X_1\|_F + \mathcal{O}(\varepsilon_M^2).$$

Thus, from the definition of $f_{TR}(\cdot)$ in (4.62) and $L_4(\cdot)$ in (4.58), $f_{TR}(m,p,p) \geq L_Y(m,p)$, leading to (4.67)–(4.68) for $k = 1$.

For $k = 2, \ldots, s$, let $t = (k-1)p$. We have that

$$\begin{aligned}\widehat{Q}_k R_k &= \begin{pmatrix} \widehat{Q}_{k-1} & Q_k \end{pmatrix} \begin{pmatrix} R_{k-1} & H_k \\ 0 & R_{kk} \end{pmatrix} \\ &= \begin{pmatrix} \widehat{X}_{k-1} + \delta \widehat{X}_{k-1} & \widehat{Q}_{k-1} H_k + Q_k R_{kk} \end{pmatrix}.\end{aligned}$$

From (A.4), (4.49), and the induction hypothesis, we have

$$\widehat{Q}_k R_k = \widehat{X}_k + \delta \widehat{X}_k,$$

where

$$\delta \widehat{X}_k = \begin{pmatrix} \delta \widehat{X}_{k-1} & \Delta Y_k - \delta Y_k \end{pmatrix}.$$

We note that

$$\|\delta \widehat{X}_k\|_F \leq \| \begin{pmatrix} \|\delta \widehat{X}_{k-1}\|_F & \|\Delta Y_k\|_F + \|\delta Y_k\|_F \end{pmatrix} \|_F.$$

Using the induction hypothesis, (A.4), and (4.49), we have

$$\begin{aligned}\|\delta \widehat{X}_k\|_F &\leq \varepsilon_M \| \begin{pmatrix} f_X(m,t,p) & L_Y(m,p) + L_1(m,p) \end{pmatrix} \|_F \|\widehat{X}_k\|_F + \mathcal{O}(\varepsilon_M^2) \\ &\leq \varepsilon_M f_X(m, kp, p) \|\widehat{X}_k\|_2 + \mathcal{O}(\varepsilon_M^2),\end{aligned}$$

which is (4.67)–(4.68). If we take $k = s$, we have (4.5).

To prove (4.69), we do another induction argument. For $k = 1$, the assumption (4.50) gives us

$$\|(I - T_1)R_1\|_F \leq \varepsilon_M L_{TR}(m,p)\|\widehat{X}_1\|_F + \mathcal{O}(\varepsilon_M^2).$$

From (4.58), we have $L_4(m,0,p) \geq L_{TR}(m,p)$; thus we have the bound (4.69).

For the induction step, we have that

$$\begin{aligned}(I - T_k)R_k &= \begin{pmatrix} I - T_{k-1} & -G_k \\ 0 & I - T_{kk} \end{pmatrix} \begin{pmatrix} R_{k-1} & H_k \\ 0 & R_{kk} \end{pmatrix} \\ &= \begin{pmatrix} (I - T_{k-1})R_{k-1} & (I - T_{k-1})H_k - G_k R_{kk} \\ 0 & (I - T_{kk})R_{kk} \end{pmatrix}.\end{aligned}$$

Thus,

$$\begin{aligned}\|(I - T_k)R_k\|_F &= \|\Gamma_k\|_F \\ &= \left\| \begin{pmatrix} \|\Gamma_{k-1}\|_F & \|(I - T_{k-1})H_k - G_k R_{kk}\|_F \\ 0 & \|\Gamma_{kk}\|_F \end{pmatrix} \right\|_F.\end{aligned}$$

Thus, we can bound the $(1,2)$ block of the above matrix from Lemma A.3 as (A.13). We use the induction hypothesis to bound the $(1,1)$ block and use assumption (4.50) to bound the $(2,2)$ block. Doing this yields

$$\begin{aligned}\|\Gamma_k\|_F &\leq \varepsilon_M \left\| \begin{pmatrix} f_{TR}(m,t,p) & L_4(m,t,p) \\ 0 & L_{TR}(m,p) \end{pmatrix} \right\|_F \|\widehat{X}_k\|_F + \mathcal{O}(\varepsilon_M^2) \\ &\leq \varepsilon_M f_{TR}(m, kp, p)\|\widehat{X}_k\|_F + \mathcal{O}(\varepsilon_M^2),\end{aligned}$$

where $f_{TR}(\cdot)$ is defined by (4.62). If we let $k = s = n/p$, we have the bound (4.6). ∎

Acknowledgments. The author acknowledges helpful conversations about topics related to this work with Ichitaro Yamazaki and Kamesh Madduri. He also acknowledges helpful suggestions from the referees.

REFERENCES

[1] N. Abdelmalek, *Roundoff error analysis for Gram–Schmidt method and solution of linear least squares problems*, BIT, 11 (1971), pp. 354–367.
[2] G. Ballard, J. Demmel, L. Grigori, M. Jacquelin, N. Knight, and H. Nguyen, *Reconstructing Householder vectors from tall-skinny QR*, J. Parallel Distrib. Comput., 85 (2015), pp. 3–31.
[3] J. Barlow, *Reorthogonalization for the Golub-Kahan-Lanczos bidiagonal reduction*, Numer. Math., 124 (2013), pp. 237–278.
[4] J. Barlow, *Block Gram-Schmidt downdating*, Electron. Trans. Numer. Anal., 43 (2014), pp. 163–187.
[5] J. Barlow, N. Bosner, and Z. Drmač, *A new backward stable bidiagonal reduction method*, Linear Algebra Appl., 397 (2005), pp. 35–84.
[6] J. Barlow and A. Smoktunowicz, *Reorthogonalized block classical Gram-Schmidt*, Numer. Math., 123 (2013), pp. 395–423.
[7] C. Bischof and C. Van Loan, *The WY representation for products of Householder matrices*, SIAM J. Sci. Stat. Comput., 8 (1987), pp. s2–s13, https://doi.org/10.1137/0908009.
[8] A. Björck, *Solving linear least squares problems by Gram–Schmidt orthogonalization*, BIT, 7 (1967), pp. 1–21.
[9] A. Björck, *Numerics of Gram–Schmidt orthogonalization*, Linear Algebra Appl., 197/198 (1994), pp. 297–316.
[10] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996, https://doi.org/10.1137/1.9781611971484.
[11] A. Björck and C. C. Paige, *Loss and recapture of orthogonality in the modified Gram–Schmidt algorithm*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 176–190, https://doi.org/10.1137/0613015.
[12] N. Bosner and J. L. Barlow, *Block and parallel versions of one-sided bidiagonalization*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 927–953, https://doi.org/10.1137/050636723.
[13] E. Carson and J. W. Demmel, *Accuracy of the s-step Lanczos method for the symmetric eigenproblem in finite precision*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 793–819, https://doi.org/10.1137/140990735.
[14] J. Demmel, L. Grigori, M. Hoemmen, and J. Langou, *Communication-optimal parallel and sequential QR and LU factorizations*, SIAM J. Sci. Comput., 34 (2012), pp. A206–A239, https://doi.org/10.1137/080731992.
[15] T. Dobravec and P. Bulić, *Comparing CPU and GPU implementations of a simple matrix multiplication algorithm*, Int. J. Comput. Elec. Engrg., 9 (2017), pp. 430–438.
[16] J. Dongarra, J. DuCroz, I. Duff, and S. Hammarling, *A set of level* 3 *basic linear algebra subprograms*, ACM Trans. Math. Software, 16 (1990), pp. 1–17.
[17] K. Fatalian, J. Sugerman, and P. Hanrahan, *Understanding the efficiency of GPU algorithms for matrix-matrix multiplication*, in Graphics Hardware 2004, T. Akenine-Möller and M. McCool, eds., ACM, New York, 2004, pp. 133–137.
[18] L. Giraud, J. Langou, M. Rozložnik, and J. V. D. Eshof, *Rounding error analysis of the classical Gram–Schmidt orthogonalization process*, Numer. Math., 101 (2005), pp. 87–100.
[19] G. Golub and C. Van Loan, *Matrix Computations*, 4th ed., The Johns Hopkins Press, Baltimore, MD, 2013.
[20] G. Golub, R. Plemmons, and A. Sameh, *Parallel block schemes for large-scale least-squares computations*, in High-Speed Computing: Scientific Applications and Algorithm Design, R. Wilhelmson, ed., University of Illinois Press, Champaign, IL, pp. 171–179.
[21] G. Golub and R. Underwood, *The block Lanczos method for computing eigenvalues*, in Mathematical Software III, J. Rice, ed., Academic Press, New York, 1977, pp. 364–377.
[22] M. Gustafsson, J. Demmel, and S. Holgren, *Numerical Evaluation of the Communication-Avoiding Lanczos Algorithm*, Technical Report, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2012.
[23] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002, https://doi.org/10.1137/1.9780898718027.
[24] M. Hoemmen, *Communication–Avoiding Krylov Subspace Methods*, Ph.D. thesis, University of California, Berkeley, CA, 2010.

[25] A. Householder, *Unitary triangularization of a nonsymmetric matrix*, J. Assoc. Comput. Mach., 5 (1958), pp. 339–342.
[26] W. Jalby and B. Philippe, *Stability analysis and improvement of the block Gram–Schmidt algorithm*, SIAM J. Sci. Stat. Comput., 12 (1991), pp. 1058–1073, https://doi.org/10.1137/0912056.
[27] C. Lawson, R. Hanson, D. Kincaid, and F. Krogh, *Basic linear algebra subprogams for FORTRAN usage*, ACM Trans. Math. Software, 5 (1979), pp. 308–323.
[28] C. C. Paige, *A useful form of unitary matrix from any sequence of unit 2-norm n-vectors*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 565–583, https://doi.org/10.1137/080725167.
[29] C. C. Paige, M. Rozložník, and Z. Strakoš, *Modified Gram-Schmidt (MGS), least squares, and backward stability of MGS-GMRES*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 264–284, https://doi.org/10.1137/050630416.
[30] C. Puglisi, *Modification of the Householder method based on the compact WY representation*, SIAM J. Sci. Stat. Comput., 13 (1992), pp. 723–726, https://doi.org/10.1137/0913042.
[31] J. Rice, *Experiments on Gram-Schmidt orthogonalization*, Math. Comp., 20 (1966), pp. 325–328.
[32] R. Schreiber and C. Van Loan, *A storage-efficient WY representation for products of Householder transformations*, SIAM J. Sci. Stat. Comput., 10 (1989), pp. 53–57, https://doi.org/10.1137/0910005.
[33] V. Simoncini and E. Gallopoulos, *Convergence properties of block GMRES and matrix polynomials*, Linear Algebra Appl., 247 (1996), pp. 97–119.
[34] V. Simoncini and E. Gallopoulos, *A hybrid block GMRES method for nonsymmetric systems with multiple right-hand sides*, J. Comput. Appl. Math., 66 (1996), pp. 457–469.
[35] K. Soodhalter, *Stagnation of block GMRES and its relationship to block FOM*, Electron. Trans. Numer. Anal., 46 (2017), pp. 162–189.
[36] A. Stathopoulos and K. Wu, *A block orthogonalization procedure with constant synchronization requirements*, SIAM J. Sci. Comput., 23 (2002), pp. 2165–2182, https://doi.org/10.1137/S1064827500370883.
[37] G. W. Stewart, *The efficient generation of random orthogonal matrices with an application to condition estimators*, SIAM J. Numer. Anal., 17 (1980), pp. 403–409, https://doi.org/10.1137/0717034.
[38] G. W. Stewart, *Block Gram–Schmidt orthogonalization*, SIAM J. Sci. Comput., 31 (2008), pp. 761–775, https://doi.org/10.1137/070682563.
[39] D. Vanderstraeten, *An accurate parallel block Gram–Schmidt algorithm without reorthogonalization*, Numer. Linear Algebra Appl., 7 (2000), pp. 219–236.
[40] H. F. Walker, *Implementation of the GMRES method using Householder transformations*, SIAM J. Sci. Stat. Comput., 9 (1988), pp. 152–163, https://doi.org/10.1137/0909010.
[41] I. Yamazaki, S. Tomov, and J. Dongarra, *Mixed-precision Cholesky QR factorization and its case studies on multicore CPU with multiple GPUs*, SIAM J. Sci. Comput., 37 (2015), pp. C307–C330, https://doi.org/10.1137/14M0973773.
[42] I. Yamazaki and K. Wu, *A communication-avoiding thick-restart Lanczos method on a distributed-memory system*, in Euro-Par 2011: Parallel Processing Workshops, M. Alexander, ed., Lecture Notes in Comput. Sci. 7155, Springer, Berlin, 2012, pp. 345–354.