Generalized QR Factorization and Its Applications*

E. Anderson
Cray Research Inc.
655F Lone Oak Drive
Eagan, Minnesota 55121

Z. Bai
Department of Mathematics
University of Kentucky
Lexington, Kentucky 40506

and

J. Dongarra
Department of Computer Science
The University of Tennessee
Knoxville, Tennessee 37996
and
Mathematical Sciences Section
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831

Submitted by F. Uhlig

ABSTRACT

The purpose of this paper is to reintroduce the generalized QR factorization with or without pivoting of two matrices A and B having the same number of rows. When B is square and nonsingular, the factorization implicitly gives the orthogonal factorization of $B^{-1}A$. Continuing the work of Paige and Hammarling, we discuss the different forms of the factorization from the point of view of general-purpose software development.

*This work was supported in part by NSF grant ASC-8715728, and in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under Contract DE-AC05-84OR21400.

655 Avenue of the Americas, New York, NY 10010 0024-3795/92/$5.00

In addition, we demonstrate the applications of the GQR factorization in solving the linear equality-constrained least-squares problem and the generalized linear regression problem, and in estimating the conditioning of these problems.

1. INTRODUCTION

The QR factorization of an $n \times m$ matrix A assumes the form

$$A = QR$$

where Q is an $n \times n$ orthogonal matrix, and $R = Q^{\mathrm{T}}A$ is zero below its diagonal. If $n \geqslant m$, then $Q^{\mathrm{T}}A$ can be written in the form

$$Q^{\mathrm{T}}A = \begin{bmatrix} R_{11} \\ 0 \end{bmatrix}$$

where R_{11} is an $n \times n$ upper triangular matrix. If $n < m$, then the QR factorization of A assumes the form

$$Q^{\mathrm{T}}A = \begin{bmatrix} R_{11} & R_{12} \end{bmatrix}$$

where R_{11} is an $n \times n$ upper triangular matrix. However, in practical applications, it is more convenient to represent the factorization in this case as

$$A = \begin{bmatrix} 0 & R_{11} \end{bmatrix} Q,$$

which is known as the RQ factorization. Closely related to the QR and RQ factorizations are the QL and LQ factorizations, which are orthogonal-lower-triangular and lower-triangular-orthogonal factorizations, respectively. It is well known that the orthogonal factors of A provide information about its column and row spaces [10].

A column pivoting option in the QR factorization allows the user to detect dependencies among the columns of a matrix A. If A has rank k, then there are an orthogonal matrix Q and a permutation matrix P such that

$$Q^{\mathrm{T}}AP = \begin{array}{cc} \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} & \begin{array}{l} k \\ n-k \end{array} \\ \begin{array}{cc} k & \quad m-k \end{array} & \end{array}$$

where R_{11} is $k \times k$, upper triangular, and nonsingular [10]. Householder transformation matrices or Givens rotation matrices provide numerically stable numerical methods to compute these factorizations with or without pivoting [10]. The software for computing the QR factorization on sequential machines is available from the public linear-algebra library LINPACK [8]. Redesigned codes in block algorithm fashion that are better suited for today's high-performance architectures will be available in LAPACK [1].

The terminology *generalized QR factorization* (GQR factorization), as used by Hammarling [12] and Paige [20], refers to the orthogonal transformations that simultaneously transform an $n \times m$ matrix A and an $n \times p$ matrix B to triangular form. This decomposition corresponds to the QR factorization of $B^{-1}A$ when B is square and nonsingular. For example, if $n \geqslant m$, $n \leqslant p$, then the GQR factorization of A and B assumes the form

$$Q^{\mathrm{T}}A = \begin{bmatrix} R \\ 0 \end{bmatrix}, \qquad Q^{\mathrm{T}}BV = \begin{bmatrix} 0 & S \end{bmatrix},$$

where Q is an $n \times n$ orthogonal matrix or a nonsingular well-conditioned matrix, V is a $p \times p$ orthogonal matrix, R is $m \times m$ and upper triangular, and S is $p \times p$ and upper triangular. If B is square and nonsingular, then the QR factorization of $B^{-1}A$ is given by

$$V^{\mathrm{T}}\left(B^{-1}A\right) = \begin{bmatrix} T \\ 0 \end{bmatrix} = S^{-1}\begin{bmatrix} R \\ 0 \end{bmatrix},$$

i.e., the upper triangular part T of the QR factorization of $B^{-1}A$ can be determined by solving the triangular matrix equation

$$S_{11}T = R,$$

where S_{11} is the $m \times m$ top left corner block of the matrix S. This implicit determination of the QR factorization of $B^{-1}A$ avoids the possible numerical difficulties in forming B^{-1} or $B^{-1}A$.

Just as the QR factorization has proved to be a powerful tool in solving least-squares and related linear regression problems, so too can the GQR factorization be used to solve both the linear equality-constrained least-squares problem

$$\min_{Bx=d} \| Ax - b \|,$$

where A and B are $m \times n$ and $p \times n$ matrices, respectively, and the generalized linear regression model

$$\min_{x,\,u} u^{\mathrm{T}} u \qquad \text{subject to} \quad b = Ax + Bu,$$

where A and B are $n \times m$ and $n \times p$ matrices, respectively. Throughout this paper, $\| \cdot \|$ denotes the Euclidean vector or matrix norm. Note that in the constrained least-squares problem, n is the column dimension of both A and B, and in the generalized regression model, n is the row dimension of both A and B.

QR factorization approaches have been used for solving these problems; see Lawson and Hanson [16, Chapters 20–22] and Björck [6, Chapter 5]. We shall see that the GQR factorization of A and B provides a uniform approach to these problems. The benefit of this approach is threefold. First, it uses a single GQR factorization concept to solve these problems directly. Second, from the software-development point of view, it allows us to develop fewer subroutines that can be used for solving these problems. Third, just as the triangular factor in the QR factorization provides important information on the conditioning of the linear least-squares problem and the classical linear regression model, the triangular factors in the GQR factorization provide information on the conditioning of these generalized problems.

Our motivation for the GQR factorization is basically the same as that of Paige [20]. However, we present a more general form of the factorization that relaxes the requirements on the rank of some of the submatrices in the factored form. This modification is significant because it simplifies the development of software to compute the factorization but does not limit the class of application problems that can be solved. We also distinguish between the GQR factorization with pivoting and without pivoting and introduce a generalized RQ factorization.

The outline of this paper is as follows: In Section 2, we show how to use the existing QR factorization and its variants to construct the GQR (or GRQ) factorization without pivoting of two matrices A and B having the same number of rows. In Section 3, we add a column pivoting option to the GQR factorization. Then, in Section 4, we show the applications of the GQR factorization in solving the linear equality-constrained least-squares problem and the generalized linear model problem, and in estimating the conditioning of these problems.

2. GENERALIZED QR FACTORIZATION

In this section, we first introduce the GQR factorization of an $n \times m$ matrix A and an $n \times p$ matrix B. For the sake of exposition, we assume

$n \geqslant m$, the most frequently occurring case. Then, for the case $n < m$, we introduce the GRQ factorization of A and B.

GQR factorization. *Let A be an $n \times m$ matrix, B an $n \times p$ matrix, and assume that $n \geqslant m$. Then there are orthogonal matrices Q $(n \times m)$ and V $(p \times p)$ such that*

$$Q^{\mathrm{T}}A = R, \qquad Q^{\mathrm{T}}BV = S, \tag{1}$$

where

$$R = \begin{array}{c} \begin{bmatrix} R_{11} \\ 0 \end{bmatrix} \begin{array}{l} m \\ n-m \end{array}, \\ m \end{array}$$

with R_{11} $(m \times m)$ upper triangular, and

$$S = \begin{array}{cc} \begin{bmatrix} 0 & S_{11} \end{bmatrix} n \\ p-n \quad n \end{array} \qquad \text{if} \quad n \leqslant p$$

where the $n \times n$ matrix S_{11} is upper triangular, or

$$S = \begin{array}{c} \begin{bmatrix} S_{11} \\ S_{21} \end{bmatrix} \begin{array}{l} n-p \\ p \end{array} \\ p \end{array} \qquad \text{if} \quad n > p,$$

where the $p \times p$ matrix S_{21} is upper triangular.

Proof. The proof is easy and constructive. By the QR factorization of A we have

$$Q^{\mathrm{T}}A = \begin{array}{c} \begin{bmatrix} R_{11} \\ 0 \end{bmatrix} \begin{array}{l} m \\ n-m \end{array}. \\ m \end{array}$$

Let Q^{T} premultiply B; then the desired factorizations follow upon the RQ factorization of $Q^{\mathrm{T}}B$. If $n \leqslant p$,

$$\left(Q^{\mathrm{T}}B\right)V = \begin{array}{cc} \begin{bmatrix} 0 & S_{11} \end{bmatrix} n; \\ p-n \quad n \end{array}$$

otherwise, the RQ factorization of $Q^{\mathrm{T}}B$ has the form

$$(Q^{\mathrm{T}}B)V = \begin{array}{c} \begin{bmatrix} S_{11} \\ S_{21} \end{bmatrix} \begin{matrix} n-p \\ p \end{matrix} \\ p \end{array}.$$

■

To illustrate these decompositions we give examples for each case:

EXAMPLE. Suppose A and B are each 4×3 matrices given by

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -3 & 2 & 1 \\ 2 & 0 & -1 \\ 3 & -1 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 4 & 3 & 1 \\ -3 & 2 & -1 \\ 1 & 3 & -1 \\ 2 & 3 & 2 \end{bmatrix}.$$

Then in the GQR factorization of A and B, the computed orthogonal matrices[1] Q and V are

$$Q = \begin{bmatrix} -0.2085 & -0.8792 & 0.1562 & -0.3989 \\ 0.6255 & -0.4147 & 0.1465 & 0.6444 \\ -0.4170 & -0.2322 & -0.7665 & 0.4296 \\ -0.6255 & 0.0332 & 0.6054 & 0.4910 \end{bmatrix},$$

$$V = \begin{bmatrix} 0.7858 & 0.5275 & -0.3229 \\ -0.1847 & -0.2982 & -0.9365 \\ -0.5902 & 0.7955 & -0.1369 \end{bmatrix},$$

and R and S are

$$R = \begin{bmatrix} -4.7958 & 1.4596 & -0.8341 \\ 0 & -2.6210 & -2.7537 \\ 0 & 0 & 2.5926 \\ 0 & 0 & 0 \end{bmatrix},$$

$$S = \begin{bmatrix} -4.2220 & 3.1170 & 0.8223 \\ -4.0063 & 1.8176 & -1.7712 \\ 0 & -2.0602 & -0.4223 \\ 0 & 0 & 3.5872 \end{bmatrix}.$$

[1]In all of these examples, the computed results are presented to four decimal digits, although the computations were carried out in double precision. If the computed variables are on the order of machine precision, (i.e., 2.2204E-016), we round to zero.

To illustrate the case $n < p$, let B be given by

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ -3 & 2 & -2 & 1 & 2 \\ 2 & 3 & 4 & -2 & -1 \\ 1 & 3 & -2 & 2 & 1 \end{bmatrix};$$

then in the GQR factorization of A and B, the orthogonal matrix V is

$$V = \begin{bmatrix} 0.3375 & -0.0791 & -0.2689 & -0.6363 & 0.6345 \\ 0.8926 & 0.2044 & -0.1635 & 0.2771 & -0.2407 \\ -0.0534 & -0.5118 & -0.5794 & -0.2833 & -0.5651 \\ 0.1585 & 0.1280 & 0.6087 & -0.6280 & -0.4401 \\ -0.2478 & 0.8208 & -0.4414 & -0.2091 & -0.1626 \end{bmatrix}$$

and the matrix S is

$$S = \begin{bmatrix} 0 & -3.4311 & 2.8692 & -1.8585 & 0.1389 \\ 0 & 0 & 7.0240 & 2.1937 & 0.1571 \\ 0 & 0 & 0 & -5.9566 & 1.0776 \\ 0 & 0 & 0 & 0 & 3.9630 \end{bmatrix}.$$

Occasionally, one wishes to compute the QR factorization of $B^{-1}A$, for example, to solve the weighted least-squares problem

$$\min_x \left\| B^{-1}(Ax - b) \right\|.$$

To avoid forming B^{-1} and $B^{-1}A$, we note that the GQR factorization (1) of A and B implicitly gives the QR factorization of $B^{-1}A$:

$$V^T(B^{-1}A) = \begin{bmatrix} T \\ 0 \end{bmatrix} = S^{-1}\begin{bmatrix} R_{11} \\ 0 \end{bmatrix},$$

i.e., the upper triangular part T of the QR factorization of $B^{-1}A$ can be determined by solving the triangular system

$$S_{11}T = R_{11}$$

for T, where S_{11} is the $m \times m$ top left corner block of the matrix S. Hence, the possible numerical difficulties in using, explicitly or implicitly, the QR factorization of $B^{-1}A$ are confined to the conditioning of S_{11}.

Moreover, if we partition $V = [V_1 \ \ V_2]$, where V_1 has m columns, then

$$B^{-1}A = V_1\left(S_{11}^{-1}R_{11}\right).$$

This shows that if A is of rank m, the columns of V_1 form an orthonormal basis for the space spanned by the columns of $B^{-1}A$. The matrix $V_1V_1^{\mathrm{T}}$ is the orthogonal projection onto $\mathscr{R}(B^{-1}A)$, where $\mathscr{R}(\cdot)$ denotes the range or column space.

Another straightforward application of the GQR factorization is to find a maximal set of BB^{T}-orthonormal vectors orthogonal to $\mathscr{R}(A)$. That is, we want to find a matrix Z such that

$$Z^{\mathrm{T}}A = 0, \qquad Z^{\mathrm{T}}BB^{\mathrm{T}}Z = I.$$

Let us rewrite the decomposition (1) as

$$\begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix} A = \begin{bmatrix} R_{11} \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix} BV = \begin{bmatrix} 0 & S_{11} & S_{12} \\ 0 & 0 & S_{22} \end{bmatrix},$$

where Q is partitioned conformally with R,

$$Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix},$$

and S_{11}, S_{22} are upper triangular. Then the desired matrix Z is given by

$$Z = Q_2 S_{22}^{-\mathrm{T}}.$$

When A is an $n \times m$ matrix with $n < m$, although it still can be presented in forms similar to that of the GQR factorization of A and B, it is sometimes more useful in applications to represent the factorization as the following:

GRQ factorization. *Let A be an $n \times m$ matrix, B an $n \times p$ matrix, and assume that $n < m$. Then there are orthogonal matrices Q $(n \times n)$ and U $(m \times m)$ such that*

$$Q^{\mathrm{T}}AU = R, \qquad Q^{\mathrm{T}}B = S, \tag{2}$$

where

$$R = \begin{array}{cc} \begin{bmatrix} 0 & R_{11} \end{bmatrix} & n \\ \begin{array}{cc} m-n & \quad n \end{array} & \end{array},$$

with R_{11} upper triangular, and

$$S = \begin{array}{cc} \begin{bmatrix} S_{11} & S_{12} \end{bmatrix} & n \\ \begin{array}{cc} n & p-n \end{array} & \end{array} \qquad \text{if} \quad n \leqslant p$$

or

$$S = \begin{array}{cc} \begin{bmatrix} S_{11} \\ 0 \end{bmatrix} & \begin{array}{l} p \\ n-p \end{array} \\ p & \end{array} \qquad \text{if} \quad n > p,$$

where the $n \times n$ or $p \times p$ matrix S_{11} is upper triangular.

Proof. The proof is similar to that of the GQR factorization. Briefly, one first does the QR factorization of B ($B = QS$), then follows it by the RQ factorization of $Q^{\mathrm{T}}A$. ∎

From the GRQ factorization of A and B, we see that if B is square and nonsingular, then the RQ factorization of $B^{-1}A$ is given by

$$(B^{-1}A)U = \begin{bmatrix} 0 & T \end{bmatrix} = S^{-1}\begin{bmatrix} 0 & R_{11} \end{bmatrix}.$$

3. GENERALIZED QR FACTORIZATION WITH PIVOTING

The previous section introduced the generalized QR factorization. As in the QR factorization of a matrix, we can also incorporate pivoting into the GQR factorization to deal with ill-conditioned or rank-deficient matrices.

GQR factorization with column pivoting. *Let A be an $n \times m$ matrix and B be an $n \times p$ matrix. Then there are orthogonal matrices Q ($n \times n$) and V ($p \times p$) and a permutation matrix P such that*

$$Q^{\mathrm{T}}AP = R, \qquad Q^{\mathrm{T}}BV = S, \tag{3}$$

where

$$R = \begin{array}{cc} \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} & \begin{array}{l} q \\ k \\ n-q-k \end{array} \\ \begin{array}{cc} q & m-q \end{array} & \end{array}$$

the $q \times q$ matrix R_{11} is upper triangular and nonsingular, and either

$$S = \begin{array}{c@{}l} \begin{bmatrix} 0 & S_{11} & S_{12} \\ 0 & 0 & S_{22} \\ 0 & 0 & 0 \end{bmatrix} & \begin{matrix} q \\ k \\ n-q-k \end{matrix} \\ \begin{matrix} p-n & q & n-q, \end{matrix} & \end{array} \qquad \textit{if} \quad n \leqslant p,$$

where the $q \times q$ matrix S_{11} is upper triangular and S_{22}, if it exists, is a full-row-rank upper trapezoidal matrix, or

$$S = \begin{array}{c@{}l} \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \\ 0 & 0 \end{bmatrix} & \begin{matrix} q \\ k \\ n-q-k \end{matrix} \\ \begin{matrix} p-n+q & n-q \end{matrix} & \end{array} \qquad \text{if} \quad n > p$$

where S_{11}, if it exists, is trapezoidal with zeros in the strictly lower left triangle, and the $k \times (n-q)$ matrix S_{22} is full-row-rank upper trapezoidal. If $p < n - q$, then the first block column of S is not present.

Proof. The proof is also constructive. By the QR factorization with pivoting of A, we have

$$Q_1^{\mathrm{T}} A P = \begin{array}{c@{}l} \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} & \begin{matrix} q \\ n-q, \end{matrix} \\ \begin{matrix} q & m-q \end{matrix} & \end{array}$$

where $q = \operatorname{rank}(A)$. If $n \leqslant p$, let

$$\left(Q_1^{\mathrm{T}} B\right) V_1 = \begin{array}{c@{}l} \begin{bmatrix} 0 & S_{11} & \bar{S}_{12} \\ 0 & 0 & \bar{S}_{22} \end{bmatrix} & \begin{matrix} q \\ n-q \end{matrix} \\ \begin{matrix} p-n & q & n-q \end{matrix} & \end{array}$$

be the RQ factorization of $Q_1^{\mathrm{T}} B$. Then by the QR factorization with pivoting on the submatrix $\bar{S}_{22}$, we have

$$Q_2^{\mathrm{T}} \bar{S}_{22} P_2 = \begin{array}{c@{}l} \begin{bmatrix} S_{22} \\ 0 \end{bmatrix} & \begin{matrix} k \\ n-q-k. \end{matrix} \\ n-q & \end{array}$$

Then the result for this case follows by setting $Q = Q_1 \operatorname{diag}(I, Q_2)$ and $V = V_1 \operatorname{diag}(I, P_2)$.

If $n > p$ and $p \leqslant n - q$, let Q_1^{T} premultiply B, and denote the result as

$$Q_1^{\mathrm{T}} B = \begin{array}{c} \begin{bmatrix} \bar{S}_{11} \\ \bar{S}_{21} \end{bmatrix} \begin{array}{l} q \\ n - q \end{array}; \\ p \end{array}$$

then by the QR factorization with pivoting of $\bar{S}_{21}$, we have

$$Q_2^{\mathrm{T}} \bar{S}_{21} P_2 = \begin{array}{c} \begin{bmatrix} S_{21} \\ 0 \end{bmatrix} \begin{array}{l} k \\ n - q - k \end{array} \\ p \end{array}$$

where $k = \operatorname{rank}(\bar{S}_{21})$. The desired factorization forms are obtained by setting $Q = Q_1 \operatorname{diag}(I, Q_2)$ and $V = P_2$. Note that in this case, the first block column of S in (3) is not present.

Otherwise, if $n > p$ and $p > n - q$, then by the RQ factorization of $Q_1^{\mathrm{T}} B$ we have

$$\left(Q_1^{\mathrm{T}} B\right) V = \begin{array}{c} \begin{bmatrix} \bar{S}_{11} \\ \bar{S}_{21} \end{bmatrix} \begin{array}{l} n - p \\ p \end{array}, \\ p \end{array}$$

where $\bar{S}_{21}$ is $p \times p$ upper triangular. The conclusion for this case follows by applying the QR factorization with pivoting to the $(n - q) \times (n - q)$ submatrix $\bar{\bar{S}}_{21}$. ∎

To illustrate these decompositions, we give an example for each case.

EXAMPLE. Let A be the 4×3 matrix

$$A = \begin{bmatrix} 1 & 3 & -3 \\ 2 & 1 & -6 \\ -1 & 1 & 3 \\ 1 & -3 & -3 \end{bmatrix},$$

where rank$(A) = 2$. To illustrate the case $n \leqslant p$, let B be the 4×5 matrix

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 3 & 2 & 3 & -2 \\ 2 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Then in the GQR decomposition with column pivoting of A and B, we have

$$Q = \begin{bmatrix} -0.3780 & 0.6412 & 0.1290 & 0.6553 \\ -0.7559 & 0.1603 & -0.3616 & -0.5217 \\ 0.3780 & 0.2565 & -0.8785 & 0.1400 \\ -0.3780 & -0.7053 & -0.2844 & 0.5281 \end{bmatrix},$$

$$V = \begin{bmatrix} 0.0930 & -0.7658 & -0.2663 & 0.1663 & -0.5535 \\ -0.5906 & 0.0188 & -0.6319 & -0.5007 & 0.0283 \\ 0.7674 & 0.2360 & -0.3643 & -0.4453 & -0.1562 \\ -0.0419 & -0.3350 & 0.6002 & -0.7234 & -0.0497 \\ -0.2279 & 0.4953 & 0.1918 & -0.0009 & -0.8161 \end{bmatrix},$$

$$P = \begin{bmatrix} e_3 & e_2 & e_1 \end{bmatrix},$$

where e_i is the ith column of an identity matrix I. The matrices R and S are

$$R = \begin{bmatrix} 7.9373 & -0.3780 & -2.6458 \\ 0 & 4.4561 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$S = \begin{bmatrix} 0.0000 & 0.9277 & 1.2449 & 2.8575 & -2.3896 \\ 0 & 0 & 0 & -1.9188 & -1.0978 \\ 0 & 0 & 0 & 6.2502 & 4.6798 \\ 0 & 0 & 0 & 0 & -3.5863 \end{bmatrix}.$$

To illustrate the case $n > p$ and $p \leqslant n - q$, let B be the 4×2 matrix

$$B = \begin{bmatrix} 2 & 3 \\ 2 & 3 \\ 2 & 3 \\ 2 & 1 \end{bmatrix}.$$

Then the GQR decomposition with column pivoting of A and B gives

$$Q = \begin{bmatrix} -0.3780 & 0.6412 & 0.1290 & 0.6553 \\ -0.7559 & 0.1603 & -0.3616 & -0.5217 \\ 0.3780 & 0.2565 & -0.8785 & 0.1400 \\ -0.3780 & -0.7053 & -0.2844 & 0.5281 \end{bmatrix},$$

$$V = \begin{bmatrix} e_2 & e_1 \end{bmatrix},$$

the matrices R and P are the same as above, and

$$S = \begin{bmatrix} -2.6458 & -2.2678 \\ 2.4685 & 0.7053 \\ -3.8609 & -3.1752 \\ 0 & 0.5272 \end{bmatrix}.$$

To illustrate the case $n > p$ and $p > n - q$, let B be the 4×3 matrix

$$B = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 1 \\ 0 & 2 & 1 \\ 2 & 2 & 0 \end{bmatrix}.$$

Then the GQR decomposition with column pivoting of A and B gives

$$Q = \begin{bmatrix} -0.3780 & 0.6412 & -0.4148 & 0.5234 \\ -0.7559 & 0.1603 & 0.1622 & -0.6137 \\ 0.3780 & 0.2565 & -0.6767 & -0.5775 \\ -0.3780 & -0.7053 & -0.5863 & 0.1264 \end{bmatrix},$$

$$V = \begin{bmatrix} -0.4380 & -0.6784 & 0.5898 \\ 0.5037 & -0.7286 & -0.4641 \\ -0.7446 & -0.0938 & -0.6609 \end{bmatrix},$$

the matrices R and P are also the same as above, and the matrix S is

$$S = \begin{bmatrix} 1.2913 & 3.4804 & -0.8051 \\ -0.8783 & -1.0619 & -0.4605 \\ 0 & 3.8337 & 0.8619 \\ 0 & 0 & 1.0103 \end{bmatrix}.$$

In Paige's work [20], the submatrix S_{11} in the definition of GQR with column pivoting is said to be of full column rank. Enforcing this assumption would make the factorization difficult to compute; in general, it would require pivoting in the first r columns of B, but such pivoting could destroy the structure of A. Our computational procedure, as outlined in the proof, simply uses the conventional QR factorization to reduce the two input matrices without requiring that S_{11} have full column rank. This makes our formulation more general and also easier to implement. As shown in the above examples, the block S_{11} in the matrix S may not be of full column rank.

Finally, we note that if B is square and nonsingular, the QR factorization with column pivoting of $B^{-1}A$ is given by

$$V^T\left(B^{-1}A\right)P = S^{-1}\begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix}.$$

4. APPLICATIONS

In this section, we shall show that the GQR factorization not only provides a simpler and more efficient way to solve the linear equality-constrained least-squares problem and the generalized linear regression problem, but also provides an efficient way to assess the conditioning of these problems. Hence the GQR factorization for solving these generalized problems is just as powerful as the QR factorization is for solving least-squares and linear regression problems. In the next section, we shall briefly mention some other applications of the GQR factorization.

4.1. *Linear Equality-Constrained Least Squares*

The linear equality-constrained least-squares (LSE) problem arises in constrained surface fitting, constrained optimization, geodetic least-squares adjustment, signal processing, and other applications. The problem is stated as follows: find an n-vector x that solves

$$\min_{Bx=d} \| Ax - b \|, \tag{4}$$

where A is an $m \times n$ matrix, $m \geqslant n$, B is a $p \times n$ matrix, $p \leqslant n$, b is an n-vector, and d is a p-vector. Clearly, the LSE problem has a solution if and only if the equation $Bx = d$ is consistent. For simplicity, we shall assume that

$$\operatorname{rank}(B) = p, \tag{5}$$

i.e., B has linearly independent rows, so that $Bx = d$ is consistent for any right-hand side d. Moreover, we assume that the null spaces $\mathcal{N}(A)$ and $\mathcal{N}(B)$ of A and B intersect only trivially:

$$\mathcal{N}(A) \cap \mathcal{N}(B) = \{0\}. \tag{6}$$

Then the LSE problem has a unique solution, which we denote by x_e. We note that (6) is equivalent to the rank condition

$$\operatorname{rank}\left(\begin{bmatrix} A \\ B \end{bmatrix}\right) = n. \tag{7}$$

Several methods for solving the LSE problem are discussed in the books by Lawson and Hanson [16, Chapters 20–22] and Björck [6, Chapter 5]. For a discussion of the large sparse matrix case, see Björck [6], Van Loan [23], Barlow et al. [3], and Barlow [4]. The null-space approach via a two-step QR decomposition is one of the most general methods for dense matrices. Now this approach can be presented more easily in terms of the GQR factorization of A and B.

By the GQR factorization of B^{T} and A^{T}, we know that there are orthogonal matrices Q and U such that

$$Q^{\mathrm{T}}A^{\mathrm{T}}U = R = \begin{array}{c} \begin{bmatrix} 0 & R_{11} & R_{12} \\ 0 & 0 & R_{22} \end{bmatrix} \begin{array}{l} p \\ n-p \end{array}, \\ \begin{array}{ccc} m-n & p & n-p \end{array} \end{array} \qquad Q^{\mathrm{T}}B^{\mathrm{T}} = S = \begin{array}{c} \begin{bmatrix} S_{11} \\ 0 \end{bmatrix} \begin{array}{l} p \\ n-p \end{array}, \\ p \end{array}$$

and from the assumptions (4) and (7), we know S_{11} and R_{22} are upper triangular and nonsingular. If we partition

$$Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}, \qquad U = \begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix},$$

where Q_1 has p columns, U_1 has $m-n$ columns, and U_2 has p columns, and set

$$y = Q^{\mathrm{T}}x = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \begin{array}{l} p \\ n-p \end{array}, \qquad c = U^{\mathrm{T}}b = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \begin{array}{l} m-n \\ p \\ n-p \end{array},$$

where $y_i = Q_i^{\mathrm{T}}x$, $i = 1, 2$, and $c_i = U_i^{\mathrm{T}}b$, $i = 1, 2, 3$, then the LSE problem is transformed to

$$\min \left\| \begin{bmatrix} 0 & 0 \\ R_{11}^{\mathrm{T}} & 0 \\ R_{12}^{\mathrm{T}} & R_{22}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} - \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \right\|$$

subject to

$$\begin{bmatrix} S_{11}^{T} & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = d.$$

Hence we can compute y_1 from the equality constraint by solving the triangular system

$$S_{11}^{\mathrm{T}} y_1 = d.$$

Then the LSE problem is truncated to the ordinary linear least-squares problem

$$\min_{y_2} \left\| R_{22}^{\mathrm{T}} y_2 - \left(c_3 - R_{12}^{\mathrm{T}} y_1 \right) \right\|.$$

Since R_{22}^{T} is nonsingular and lower triangular, y_2 is given by

$$y_2 = R_{22}^{-\mathrm{T}} \left(c_3 - R_{12}^{\mathrm{T}} y_1 \right),$$

which only involves solving a triangular system. The solution of the LSE problem is then given by

$$x_e = Qy = Q_1 y_1 + Q_2 y_2,$$

or in a more straightforward form,

$$x_e = Q_2 R_{22}^{-\mathrm{T}} U_3^{\mathrm{T}} b + Q \begin{bmatrix} I \\ -R_{22}^{-\mathrm{T}} R_{12}^{\mathrm{T}} \end{bmatrix} S_{11}^{-\mathrm{T}} d,$$

and the residual sum of squares $\rho^2 = \|r_e\|^2 = \|Ax_e - b\|^2$ is given by

$$\rho^2 = \|c_1\|^2 + \|R_{11}^{\mathrm{T}} y_1 - c_2\|^2.$$

EXAMPLE. Let the LSE problem be specified with

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \qquad d = \begin{bmatrix} 7 \\ 4 \end{bmatrix}.$$

The exact solution to this problem is $x_e = \frac{1}{8}[46, -2, 12]^{\mathrm{T}}$. By the GQR factorization of B^{T} and A^{T}, we have

$$R^{\mathrm{T}} = \begin{bmatrix} 0 & 1.3583 & 3.1867 & 1.6330 \\ 0 & 0 & 0 & 1.1547 \\ 0 & 0 & 0 & -2.0000 \end{bmatrix}, \quad S^{\mathrm{T}} = \begin{bmatrix} -1.7321 & -0.5774 \\ 0 & -1.6330 \\ 0 & 0 \end{bmatrix},$$

and the computed solution of the LSE problem is

$$\hat{x}_e = \begin{bmatrix} 5.7500 \\ -0.2500 \\ 1.5000 \end{bmatrix}.$$

The relative error of the computed solution is

$$\frac{\| x_e - \hat{x}_e \|}{\| x_e \|} \approx 4.2892 \times 10^{-16},$$

and the norm of the residual $\| A\hat{x}_e - b \| \approx 9.2466$.

The Sensitivity of the LSE Problem. The condition numbers of A and B were introduced by Eldèn [9] to assess the perturbed behavior of the LSE problem. Specifically, let E be an error matrix of A, F be an error matrix of B, and e and f be errors of b and d, respectively. We assume that $B + F$ also has full row rank and $\mathcal{N}(A + E) \cap \mathcal{N}(B + F) = \{0\}$, i.e., the perturbed LSE problem also has a unique solution. Let $\bar{x}_e$ be the solution of the same problem with A, B, b, and d replaced by $A + E$, $B + F$, $b + e$, and $d + f$, respectively. Eldèn introduced the condition numbers

$$\kappa_B(A) = \| A \| \left\| (AG)^\dagger \right\|, \qquad \kappa_A(B) = \| B \| \, \| B_A^\dagger \|$$

to measure the sensitivity of the LSE problem, where

$$G = I - B^\dagger B, \qquad B_A^\dagger = \left[I - (AG)^\dagger A \right] B^\dagger,$$

and $A^\dagger$ denotes the Moore-Penrose pseudoinverse of a matrix A.

Under mild conditions, Eldèn's asymptotic perturbation bound, modified slightly here, can be presented as follows:

LSE-Problem Perturbation Bound.

$$\frac{\| x_e - \bar{x}_e \|}{\| x_e \|} \leqslant \kappa_B(A) \left(\frac{\| E \|}{\| A \|} + \nu_e \right) + \kappa_A(B) \left(\frac{\| F \|}{\| B \|} + \gamma_e \right)$$
$$+ \kappa_B^2(A) \left(\frac{\| E \|}{\| A \|} + \kappa_A(B) \frac{\| F \|}{\| B \|} \right) \rho_e + O(\varepsilon^2),$$

where

$$\nu_e = \frac{\| e \|}{\| A \| \, \| x_e \|}, \qquad \gamma_e = \frac{\| f \|}{\| B \| \, \| x_e \|}, \qquad \rho_e = \frac{\| r_e \|}{\| A \| \, \| x_e \|}, \qquad r_e = Ax_e - b,$$

and $O(\varepsilon^2)$ denotes the higher-order term in the perturbation matrices E, F, etc.

The interpretation of this result is that the sensitivity of $\bar{x}_e$ is measured by $\kappa_B(A)$ and $\kappa_A(B)$ if the residual r_e is zero or relatively small, and otherwise by $\kappa_B^2(A)[\kappa_A(B)+1]$.

We note that if the matrix B is zero (hence $F = 0$), then the LSE problem is just the ordinary linear least-squares problem. The perturbation bound for the LSE problem is then reduced to

$$\frac{\|x_e - \bar{x}_e\|}{\|x_e\|} \leqslant \kappa(A)\left(\frac{\|E\|}{\|A\|} + \frac{\|e\|}{\|A\|\,\|x_e\|}\right) + \kappa^2(A)\frac{\|E\|}{\|A\|}\frac{\|r_e\|}{\|A\|\,\|x_e\|} + O(\varepsilon^2),$$

where $\kappa_B(A) = \kappa(A) = \|A\|\,\|A^{\dagger}\|$. This is just the perturbation bound of the linear least-squares problem obtained by Golub and Wilkinson [10].

Estimation of the Condition Numbers. The condition numbers $\kappa_B(A)$ and $\kappa_A(B)$ of the LSE problem involves $B^{\dagger}$, $B^{\dagger}B$, $(AG)^{\dagger}$. etc., and computing these matrices can be expensive. Fortunately, it is possible to compute inexpensive estimates of $\kappa_B(A)$ and $\kappa_A(B)$ without forming $B^{\dagger}$, $B^{\dagger}B$, or $(AG)^{\dagger}$. This can be done using a method of Hager [11] and Higham [14] that computes a lower bound for $\|B\|_{\infty}$, where B is a matrix, given a means for evaluating matrix-vector products Bu and $B^{\mathrm{T}}u$. Typically, four or five products are required, and the lower bound is almost always within a factor 3 of $\|B\|_{\infty}$. To estimate $\kappa_B(A)$ and $\kappa_A(B)$, we need to estimate vector norms $\|Kz\|_{\infty}$, where $K = (AG)^{\dagger}$ or $K = B_A^{\dagger}$, and $z \geqslant 0$ is a vector that is readily computed. Given the GQR factorization of A and B, after tedious computations, we have

$$(\mathrm{AG})^{\dagger} z = Q_2 R_{22}^{-\mathrm{T}} U_3^{\mathrm{T}} z,$$

$$B_A^{\dagger} z = Q\begin{bmatrix} I \\ -R_{22}^{-\mathrm{T}} R_{12}^{\mathrm{T}} \end{bmatrix} S_{11}^{-\mathrm{T}} z,$$

where we do not need to form $R_{22}^{-\mathrm{T}}$ or $S_{11}^{-\mathrm{T}}$, but rather solve the triangular system and do matrix-vector operations.

Roughly speaking, the conditioning of the LSE problem only depends on the conditioning of the matrices R_{22} and S_{11}. In the last example, although the matrix R is ill conditioned (actually, it is singular), we have

$$R_{22} = [-2] \quad \text{and} \quad S_{11} = \begin{bmatrix} -1.7321 & -5.7735 \\ 0 & -1.6330 \end{bmatrix},$$

so it turns out to be a well-conditioned problem.

4.2. Generalized Linear Regression Model

The generalized linear regression model (GLM) problem can be written as

$$b = Ax + w, \tag{8}$$

where w is a random error with mean 0 and a symmetric nonnegative definite variance-covariance matrix $\sigma^2 W$. The problem is that of estimating the unknown parameters x on the basis of the observation b. If W has rank p, then W has a factorization

$$W = BB^{\mathrm{T}},$$

where the $n \times p$ matrix B has linearly independent columns (for example, the Cholesky factorization of W could be carried out to get B). In some practical problems, the matrix B might be available directly. For numerical computation reasons it is preferable to use B rather than W, since W could be ill conditioned, but the condition of B may be much better. Thus we replace (8) by

$$b = Ax + Bu, \tag{9}$$

where A is an $n \times m$ matrix, B is an $n \times p$ matrix, and u is a random error with mean 0 and covariance $\sigma^2 I$. Then the estimator of x in (9) is the solution to the following algebraic generalized linear least-squares problem:

$$\min_{x,u} u^T u \qquad \text{subject to} \quad b = Ax + Bu. \tag{10}$$

Notice that this problem is defined even if A and B are rank-deficient. For convenience, we assume that $n \geqslant m$, $n \geqslant p$, the most frequently occurring case. When $B = I$, (10) is just an ordinary linear regression problem. We assume that the matrices A and B in (10) are general dense matrices. If we know A or B has a special structure, e.g. if B is triangular, then we might need to take a different approach in order to save the work without destroying the structure (see, for example, [15]).

The GLM problem can be formulated as the LSE problem:

$$\min \left\| \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \right\| \qquad \text{subject to} \quad \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = b.$$

Hence, it is easy to see that the GLM problem has a solution if the linear system

$$\begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = b$$

is consistent. Because of high overhead and possible numerical difficulties if the matrices A and B are scaled differently, it is not advisable to solve the GLM problem directly by the method of the LSE problem. Paige [18] and Hammarling [12] proposed a two-step QR decomposition approach to the GLM problem to treat A and B separately. Now, we show that this approach can be simplified with GQR-factorization terminology.

By the GQR factorization with pivoting of A and B, we have orthogonal matrices Q $(n \times n)$ and V $(p \times p)$ and a permutation matrix P such that

$$Q^{\mathrm{T}}AP = \begin{array}{cc} \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} & \begin{array}{l} q \\ n-q \end{array} \\ \begin{array}{cc} q & m-q \end{array} & \end{array}, \qquad Q^{\mathrm{T}}BV = \begin{array}{cc} \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} & \begin{array}{l} q \\ n-q \end{array} \\ \begin{array}{cc} p-n+q & n-q \end{array} & \end{array},$$

where the $q \times q$ matrix R_{11} is upper triangular and nonsingular. We also assume that the $(n-q) \times (n-q)$ matrix S_{22} is upper triangular and nonsingular for simplicity of exposition. If we partition

$$Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}, \qquad V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}, \qquad P = \begin{bmatrix} P_1 & P_2 \end{bmatrix},$$

where Q_1 has q columns, V_2 has $n-q$ columns, and P_1 has q columns, and set

$$c = Q^{\mathrm{T}}b \equiv \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \qquad v = V^{T}u \equiv \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \qquad y = P^{\mathrm{T}}x \equiv \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

i.e., $c_i = Q_i^{\mathrm{T}}b$, $v_i = V_i^{\mathrm{T}}u$, $y_i = P_i^{\mathrm{T}}x$, $i = 1, 2$, then the constrained equation of the GLM problem (10) is transformed to

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \tag{11}$$

Hence v_2 can be determined from the "bottom" equation of (11) by solving a triangular system

$$S_{22}v_2 = c_2.$$

Then from the "top" equation of (11), we have

$$c_1 = R_{11}y_1 + R_{12}y_2 + S_{11}v_1 + S_{12}v_2.$$

It is obvious that to get the minimum-2-norm solutions, the remaining components of the solutions can be chosen as

$$v_1 = 0, \qquad y_2 = 0, \qquad y_1 = R_{11}^{-1}(c_1 - S_{12}v_2).$$

Then the solutions of the original problem are

$$x_e = P_1 R_{11}^{-1}\left(Q_1^{\mathrm{T}} - S_{12}S_{22}^{-1}Q_2^{\mathrm{T}}\right)b, \qquad u_e = V_2 S_{22}^{-1}Q_2^{\mathrm{T}}b.$$

EXAMPLE. Let the matrices A, B and the vector b in the GLM problem be

$$A = \begin{bmatrix} 1 & 2 & 1 & 4 \\ 1 & 1 & 1 & 1 \\ -1 & -2 & -1 & 1 \\ -1 & 2 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 2 & 2 \\ -1 & 1 & -2 \\ 3 & 1 & 6 \\ 2 & -2 & 4 \\ 1 & -1 & 2 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

where $\operatorname{rank}(A) = 3$, $\operatorname{rank}(B) = 2$. The exact solutions of the GLM problem are $x_e = \frac{1}{9}[0, 6, 10, -16]^{\mathrm{T}}$ and $u_e = \frac{1}{45}[14, 70, 28]^{\mathrm{T}}$.

By the GQR factorization with column pivoting of the matrices A and B, we have

$$R = \begin{bmatrix} -4.4721 & -1.3416 & -1.3416 & -1.3416 \\ 0 & -3.4928 & -6.2986 & -6.2986 \\ 0 & 0 & 1.6743 & 1.6743 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$S = \begin{bmatrix} 0 & 1.6676 & 3.0853 \\ 0 & -0.4724 & -1.7320 \\ 0 & 6.7913 & 1.6763 \\ 0 & -5.1167 & -0.0329 \\ 0 & 0 & 0.6015 \end{bmatrix}.$$

Then the computed solutions are

$$\hat{x}_e = \begin{bmatrix} 0 \\ 0.6667 \\ 2.1111 \\ -1.7778 \end{bmatrix}, \qquad \hat{u}_e = \begin{bmatrix} 0.3111 \\ 1.5556 \\ 0.6222 \end{bmatrix}.$$

The relative errors of the computed solutions are

$$\frac{\| x_e - \hat{x}_e \|_2}{\| x_e \|_2} \approx 7.9752 \times 10^{-16}, \qquad \frac{\| u_e - \hat{u}_e \|_2}{\| u_e \|_2} \approx 6.6762 \times 10^{-16}.$$

The square minimal length of the vector $\hat{u}_e$ is $\hat{u}_e^{\mathrm{T}} \hat{u}_e \approx 2.9037$, and the residual is $\| b - A\hat{x}_e - B\hat{u}_e \|_2 \approx 4.4464 \times 10^{-15}$.

The Sensitivity of the GLM Problem. Regarding the sensitivity of the problem to perturbations, we shall consider the effects of the perturbations in the vector b and in the matrices A and B. Let the perturbed GLM problem be defined as

$$\min_{\bar{x}, \bar{u}} \bar{u}^{\mathrm{T}} \bar{u} \qquad \text{subject to} \quad b + e = (A + E)\bar{x} + (B + F)\bar{u}.$$

The solutions are denoted by $\bar{x}_e$ and $\bar{u}_e$. Then under the assumptions

$$\operatorname{rank}(A) = \operatorname{rank}(A + E) = m$$

and

$$\operatorname{rank}(A, B) = \operatorname{rank}(A + E, B + F) = n,$$

we have the following bounds on the relative error in $\bar{x}_e$ and $\bar{u}_e$ due to the perturbations of b, A, and B:

GLM-Problem Perturbation Bounds.

$$\frac{\| \bar{x}_e - x_e \|}{\| x_e \|} \leqslant \kappa_B(A) \left(\frac{\| E \|}{\| A \|} + \frac{\| e \|}{\| A \| \, \| x_e \|} \right) + \kappa_B(A) \left(\kappa_B(A) \frac{\| E \|}{\| A \|} + \frac{\| F_1 \|}{\| B \|^2} \right) \frac{\| B^2 \| \, \| p \|}{\| A \| \, \| x_e \|} + O(\varepsilon^2) \tag{12}$$

and

$$\frac{\| \bar{u}_e - u_e \|}{\| b \|} \leqslant \kappa_B(A) \frac{\| E \|}{\| A \|} \frac{\| B \| \, \| p \|}{\| b \|} + \frac{\kappa_A(B)}{\| B \|} \left(\| E \| \frac{\| x_e \|}{\| b \|} + \frac{\| e \|}{\| b \|} + \| F_1 \| \frac{\| p \|}{\| b \|} \right) + \| F \| \frac{\| p \|}{\| b \|} + O(\varepsilon^2), \tag{13}$$

where

$$\kappa_B(A) = \|A\|\,\|A_B^\dagger\|, \qquad \kappa_A(B) = \|B\|\,\|(GB)^\dagger\|,$$

$G = I - AA^\dagger$, $A_B^\dagger = A^\dagger[I - B(GB)^\dagger]$, and $p = [GB(GB)^{\mathrm{T}}]^\dagger b$, $F_1 = FB^{\mathrm{T}} + BF^{\mathrm{T}}$. Here $O(\varepsilon^2)$ represents the higher-order term in the perturbation matrices E, F, etc.

The proof is long and appears in the appendix.

If we note that

$$\|B\|^2\|p\| \leqslant \kappa_A^2(B)\|b\|,$$

then the bounds (12) and (13) can be simplified. We see that the sensitivities of $\bar{x}_e$ and $\bar{u}_e$ basically depend on $\kappa_B(A)$ and $\kappa_A(B)$. For this reason, $\kappa_B(A)$ and $\kappa_A(B)$ are defined as the *condition numbers* of the GLM problem. They can be used to predict the effects of errors in the regression variables on regression coefficients.

As a special case, we note that if $B = I$, then the GLM problem is reduced to the classical linear regression problem. Then u_e is just the residual vector, $u_e = r_e = b - Ax_e$, $\bar{u}_e = \bar{r}_e = (b + e) - (A + E)\bar{x}_e$, $F = 0$, and

$$\kappa_B(A) = \kappa(A) = \|A\|\,\|A^\dagger\|, \qquad \kappa_A(B) = 1.$$

Hence we have

$$\frac{\|\bar{x}_e - x_e\|}{\|x\|} \leqslant \kappa(A)\left(\frac{\|E\|}{\|A\|} + \frac{\|e\|}{\|A\|\,\|x_e\|}\right) + \kappa^2(A)\frac{\|E\|}{\|A\|}\frac{\|r_e\|}{\|A\|\,\|x_e\|} + O(\varepsilon^2)$$

and

$$\frac{\|\bar{r}_e - r_e\|}{\|b\|} \leqslant \kappa(A)\frac{\|E\|}{\|A\|}\frac{\|r_e\|}{\|b\|} + \|E\|\frac{\|x_e\|}{\|b\|} + \frac{\|e\|}{\|b\|} + O(\varepsilon^2)$$

These are the well-known perturbation results for the solution and residual of the ordinary linear regression problem [22, 10].

Estimation of Condition Numbers. To estimate the condition numbers $\kappa_B(A)$ and $\kappa_A(B)$ of the GLM problem, we again can use the Hager-Higham method without the expense of forming $A^\dagger$ or $(GB)^\dagger$. By this technique, the required vector norms $\|Kz\|_\infty$ can be computed from the GQR factorization of A and B, where $K = (GB)^{\dagger}$ or $K = A_B^{\dagger}$ and $z \geqslant 0$ is a vector that is readily computed. After tedious computations, we have

$$(GB)^{\dagger} z = V_2 S_{22}^{-1} Q_2^{\mathrm{T}} z$$

$$A_B^{\dagger} z = P_1 R_{11}^{-1} \left(Q_1^{\mathrm{T}} z - S_{12} S_{22}^{-1} Q_2^{\mathrm{T}} z \right).$$

Hence, we can just use a triangular system solver and matrix-vector operations to give the estimation of condition numbers of the GLM problem.

Roughly speaking, we see that the conditioning of the GLM problem depends on the conditioning of the triangular matrices R_{11} and S_{22}.

4.3. *Other Applications*

In this section, we briefly mention some other applications of the GQR factorizations.

The GQR factorization has been used as a preprocessing step for computing the generalized singular-value decomposition in the Jacobi-Kogbetliantz approach; see Paige [19] and Bai [2].

The GQR factorization can also be used in solving structural equations:

$$f = A^{\mathrm{T}} t, \qquad e = BB^{\mathrm{T}} t, \qquad e = -Ad,$$

where f is given, and we wish to find d. This kind of problem regularly arises in the analysis of structures made up of elements joined in the style of a framework or network; see Heath et al. [13] and Paige [20].

5. SUMMARY AND FUTURE WORK

In this paper, we have defined the generalized QR factorization with or without partial pivoting of two matrices A and B, each having the same number of rows, and shown its applications in solving the linear equality-constrained least-squares problem and generalized linear model problems, and in assessing the conditioning of these problems. A similar development could be done for matrices A and B having the same number of columns, instead of the same number of rows. Then the GQR factorization of A and B would be equivalent to the QR factorization of AB^{-1}. These discussions have served as the guideline for our future development of GQR factorization software for the LAPACK library [1].

APPENDIX

In this appendix, we prove the perturbation bounds (12) and (13) for the solutions x and u of the GLM problem presented in Section 4.2.

The Lagrangian of the GLM problem is

$$\lambda(x, u, p) = u^{\mathrm{T}}u + 2p^{\mathrm{T}}(b - Ax - Bu),$$

where p is a vector of Lagrange multipliers. Taking derivatives with respect to x, u, and p and equating the results to zero gives the first-order necessary conditions for the minimum:

$$\begin{bmatrix} 0 & 0 & A^{\mathrm{T}} \\ 0 & -I & B^{\mathrm{T}} \\ A & B & 0 \end{bmatrix} \begin{bmatrix} x \\ u \\ p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ b \end{bmatrix}. \tag{14}$$

(The theory may be found in most textbooks dealing with constrained optimization; see for example [17].) Since this is a linear equality-constrained problem and the Hessian of the objective function is $2I$, which always is positive definite, any solution of (14) also solves the GLM problem, so that (14) is necessary and sufficient for the GLM problem. Here we can eliminate $u = B^{\mathrm{T}}p$ to give

$$\begin{bmatrix} 0 & A^{\mathrm{T}} \\ A & -BB^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} x \\ -p \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix}. \tag{15}$$

Similarly, the perturbed GLM problem can be reformulated as

$$\begin{bmatrix} 0 & A^{\mathrm{T}} + E^{\mathrm{T}} \\ A + E & -(B + F)(B + F)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} x + \Delta x \\ -(p + \Delta p) \end{bmatrix} = \begin{bmatrix} 0 \\ b + e \end{bmatrix}. \tag{16}$$

The (pseudo)inverse of the coefficient matrix is in the following lemma, which is due to Eldèn [9]; we have modified it slightly to fit our case.

LEMMA. *Let*

$$C = \begin{bmatrix} 0 & A^{\mathrm{T}} \\ A & -BB^{\mathrm{T}} \end{bmatrix}$$

and

$$Y = \begin{bmatrix} \left(A_B^\dagger B\right)\left(A_B^\dagger B\right)^{\mathrm{T}} & A_B^\dagger \\ \left(A_B^\dagger\right)^{\mathrm{T}} & G^{\mathrm{T}}\left[(GB)(GB)^{\mathrm{T}}\right]^\dagger G \end{bmatrix},$$

where

$$G = I - AA^\dagger, \qquad A_B^\dagger = A^\dagger\left[I - B\left(GB\right)^\dagger\right].$$

If $rank([A, B]) = n$, *then* $Y = C^\dagger$. *Further, if* A *has full column rank, then* $Y = C^{-1}$.

Subtracting the matrix equation (15) from (16), we have

$$\begin{bmatrix} \Delta x \\ \Delta p \end{bmatrix} = -\left(C + E_c\right)^{-1} E_c \begin{bmatrix} x \\ p \end{bmatrix} + \left(C + E_c\right)^{-1} \begin{bmatrix} 0 \\ e \end{bmatrix}, \tag{17}$$

where

$$E_c = \begin{bmatrix} 0 & E^{\mathrm{T}} \\ E & -F_1 \end{bmatrix}, \qquad F_1 = FB^{\mathrm{T}} + BF^{\mathrm{T}}.$$

If $\|C^{-1}E_c\| < 1$, then we can make the expansion

$$\left(C + E_c\right)^{-1} = C^{-1} - C^{-1}E_cC^{-1} + \cdots,$$

and then (17) becomes

$$\begin{bmatrix} \Delta x \\ \Delta p \end{bmatrix} = -C^{-1}E_c \begin{bmatrix} x \\ p \end{bmatrix} + C^{-1}\begin{bmatrix} 0 \\ e \end{bmatrix} + O\left(\varepsilon^2\right),$$

where $O(\varepsilon^2)$ means the higher-order terms in the perturbation factors E_c and e, which we omit in the formulas that follow. By the lemma, we have

$$\Delta x = -\left(A_B^\dagger B\right)\left(A_B^\dagger B\right)^{\mathrm{T}} E^{\mathrm{T}} p - A_B^\dagger\left(Ex + F_1 p - e\right).$$

After taking norms, it becomes

$$\|\Delta x\| \leqslant \|B\|^2 \|A_B^\dagger\|^2 \|E\| \, \|p\| + \|A_B^\dagger\|\left(\|E\| \, \|x\| + \|F_1\| \, \|p\| + \|e\|\right).$$

Using the condition numbers

$$\kappa_B(A) = \|A\| \, \|A_B^\dagger\|, \qquad \kappa_A(B) = \|B\| \, \|(GB)^\dagger\|,$$

we get the desired relative perturbation bound (12) on the solution x.

For the perturbation bound on the solution u, from (17), we first have

$$\Delta p = -\left(A_B^\dagger\right)^{\mathrm{T}} E^{\mathrm{T}} p + G^{\mathrm{T}}\left[(GB)(GB)^T\right]^\dagger G\left(Ex + F_1 p + e\right).$$

Since $u = B^{\mathrm{T}}p$ and $u + \Delta u = (B + F)^{\mathrm{T}}(p + \Delta p)$, we have, subtracting them and using $A^\dagger = A^{\mathrm{T}}(AA^{\mathrm{T}})^\dagger$,

$$\begin{aligned}\Delta u &= B^{\mathrm{T}} \Delta p + Fp \\ &= B^{\mathrm{T}}\left(A_B^\dagger\right)^{\mathrm{T}} E^{\mathrm{T}} p + B^{\mathrm{T}} G^{\mathrm{T}}\left[(GB)(GB)^{\mathrm{T}}\right]^\dagger G\left(Ex + F_1 p + e\right) \\ &= B^{\mathrm{T}}\left(A_B^\dagger\right)^{\mathrm{T}} E^{\mathrm{T}} p + \left(GB\right)^\dagger G\left(Ex + F_1 p + e\right),\end{aligned}$$

where the higher-order terms of the perturbation factors of E, F, and e again are not presented. By taking norms, and substituting in the condition numbers $\kappa_B(A)$ and $\kappa_A(B)$, we get the desired perturbation bound (13) on the solution u.

The authors are very grateful to Jim Demmel, Sven Hammarling, and Jeremy Du Croz for their valuable comments. The authors would also like to express special thanks to the referees, whose comments led to improvements in the presentation.

REFERENCES

1 E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, LAPACK: A portable linear algebra library for high performance computers, in *Proceedings Supercomputing '90*, IEEE Computer Soc. Press, Los Alamitos, Calif., 1990, pp. 2–11.

2 Z. Bai, An improved algorithm for computing the generalized singular value decomposition, *Numer. Math.*, submitted for publication.

3 J. L. Barlow, N. K. Nichols, and R. J. Plemmons, Iterative methods for equality constrained least squares problems, *SIAM J. Sci. Statist. Comput.* 9:892–906 (1988).

4 J. L. Barlow, Error analysis and implementation aspects of deferred correction for equality constrained least squares problems, *SIAM J. Numer. Anal.* 25:1340–1358 (1989).

5 C. H. Bischof and P. C. Hansen, Structure-Preserving and Rank-Revealing QR-Factorizations, Preprint MCS-P100-0989, Argonne National Lab., 1989.

6 Åke Björck, Least squares methods, in *Handbook of Numerical Analysis. Vol. 1: Solution of Equations in* R^n (P. G. Ciarlet and J. L. Lions, Eds.), Elsevier North Holland, 1987.

7 T. F. Chan, Rank revealing *QR* factorizations, *Linear Algebra Appl.* 88/89:67–82 (1987).

8 J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK User's Guide*, SIAM, Philadelphia, 1979.

9 L. Eldèn, Perturbation theory for the least squares problem with linear equality constraints, *SIAM J. Numer. Anal.* 17:338–350 (1980).

10 G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins U.P., Baltimore, 1989.

11 W. W. Hager, Condition Estimates, *SIAM J. Sci. Statist. Comput.* 31(2):221–239 (1984).

12 S. Hammarling, The numerical solution of the general Gauss-Markov linear model, in *Mathematics in Signal Processing* (T. S. Durrani et al., Eds.), Clarendon Press, Oxford, 1986.

13 M. T. Heath, R. J. Plemmons, and R. C. Ward, Sparse orthogonal schemes for structural optimization using the force method, *SIAM J. Sci. Statist. Comput.* 7:1147–1159 (1984).

14 N. J. Higham, FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation, *ACM Trans. Math. Software* 14:381–396 (1988).

15 S. Kourouklis and C. Paige, A constrained least squares approach to the general Gauss-Markov model, *J. Amer. Statist. Assoc.* 76:620–625 (1981).

16 C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N.J., 1974.

17 D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed., Addison-Wesley, Reading, Mass., 1984.

18 C. Paige, Computer solution and perturbation analysis of generalized linear least squares problems, *Math. Comput.* 33:171–183 (1979).

19 C. Paige, Computing the generalized singular value decomposition, *SIAM J. Sci. Statist. Comput.* 7:1126–1146 (1986).

20 C. Paige, Some aspects of generalized QR factorization, in *Reliable Numerical Computations* (M. Cox and S. Hammarling, Eds.), Clarendon Press, Oxford, 1990.

21 R. Schreiber and C. Van Loan, A storage efficient *WY* representation for products of Householder transformations, *SIAM J. Sci. Statist. Comput.* 10(1):53–57 (Jan. 1989).
22 G. W. Stewart, On the perturbation of pseudo-inverses, projections and linear least squares problems, *SIAM Rev.* 19:634–662 (1977).
23 C. F. Van Loan, On the method of weighting for equality-constrained least-squares problems, *SIAM J. Numer. Anal.* 22:851–864 (1985).

Received 31 July 1990; final manuscript accepted 6 March 1991