MATHEMATICS OF COMPUTATION
VOLUME 43, NUMBER 168
OCTOBER 1984, PAGES 483-489

On the Asymptotic Behavior of Scaled Singular Value and QR Decompositions

By G. W. Stewart*

Abstract. Asymptotic expressions are derived for the singular value decompositon of a matrix some of whose columns approach zero. Expressions are also derived for the QR factorization of a matrix some of whose rows approach zero. The expressions give insight into the method of weights for approximating the solutions of constrained least squares problems.

1. Introduction. It is well known that certain widely used matrix decompositions change in nontrivial ways when their rows or columns are multiplied by constants. For example, let the $n \times p$ matrix X be partitioned in the form

$$X = (X_1\ X_2), \tag{1.1}$$

and for $1 \geqslant t \geqslant 0$ define**

$$X_t = (X_1\ tX_2). \tag{1.2}$$

Let

$$X_t = U_t S_t V_t^T \tag{1.3}$$

be the singular value decomposition of X_t (see [4] for definitions). The columns of U_t and V_t (the singular vectors of X_t) and the diagonal elements of S_t (the singular values) are nonlinear functions of t, and there is no simple way of obtaining, say, S_t from S_1.

One purpose of this paper is to derive expansions for the singular value decomposition of X_t as t approaches zero. An application of these expansions is the following. When t is small, the ratio of the largest to the smallest singular values of X_t will be large, and from this one might conclude that problems associated with X_t are ill-conditioned. In order to determine whether this apparent ill-conditioning is real, it is necessary to have precise information about the singular value decomposition of X_t (see [3] for an example).

A related problem concerns the QR factorization of a matrix. Partition X in the form

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \tag{1.4}$$

Received August 31, 1983.

1980 *Mathematics Subject Classification*. Primary 15A18, 15A21, 65F35, 62J99.

*This work was done while the author was at the Scientific Computing Division of the Center for Applied Mathematics of the National Bureau of Standards.

**There is an ambiguity in this notation, since X_1 can mean either the first element in the partition (1.1) or the matrix X_t for $t = 1$. To resolve it, we let X_1 refer only to the former.

where X_{11} is square, and write

$$X_t = \begin{bmatrix} X_{11} & X_{12} \\ tX_{21} & tX_{22} \end{bmatrix} \tag{1.5}$$

[note the different definitions of X_t in (1.2) and (1.5)]. Matrices such as (1.5) with small t arise in the numerical solution of constrained least squares problems [1, Chapter 22]. The approach is to down-weight the problem [represented by $(X_{21}\ X_{22})$] compared to the constraints [represented by $(X_{11}\ X_{12})$]. The best method for solving constrained problems in this way uses the QR factorization $X_t = Q_t R_t$, in which Q_t has orthonormal columns and R_t is upper triangular. In Section 3 we will derive asymptotic expressions for Q_t and R_t.

2. The Singular Value Decomposition. We begin by observing that null vectors of X behave in a simple manner under scaling. If $Xv = 0$ and

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

is partitioned conformally with (1.1), then

$$v_t = \begin{bmatrix} tv_1 \\ v_2 \end{bmatrix}$$

is a null vector of X_t. Thus null vectors behave linearly under transformations of the form (1.2), and there is no need to treat them here. We shall therefore assume that X has no null vectors.

The main result is summarized in the following theorem.

THEOREM 2.1. *Let the $n \times p$ matrix X have rank p, and let X be partitioned as in* (1.1). *Let X_t be defined by* (1.2). *Let*

$$B = \left(X_1^T X_1\right)^{-1} X_1^T X_2, \tag{2.1}$$

and

$$\overline{X}_2 = X_2 - X_1 B. \tag{2.2}$$

To each singular value s_1 of X_1 there is associated a unique singular value $s_1^{(t)}$ of X_t which satisfies

$$s_1^{(t)} = s_1 + O(t^2). \tag{2.3}$$

If s_1 is simple and its right singular vector is denoted by v_1, then the corresponding right singular vector of X_t satisfies

$$v_1^{(t)} = \begin{bmatrix} v_1 + O(t^2) \\ tB^T v_1 + O(t^3) \end{bmatrix}. \tag{2.4}$$

Moreover, if the left singular vector of s_1 is denoted by u_1, then the corresponding left singular vector of X_t satisfies

$$u_1^{(t)} = u_1 + O(t^2). \tag{2.5}$$

To each singular value $\bar{s}_2$ of $\overline{X}_2$ there is associated a unique singular value $s_2^{(t)}$ of X_t which satisfies

$$s_2^{(t)} = t\bar{s}_2 + O(t^3). \tag{2.6}$$

If $\bar{s}_2$ is simple and its right singular vector is denoted by $\bar{v}_2$, then the corresponding right singular vector of X_t satisfies

$$v_2^{(t)} = \begin{bmatrix} -tB\bar{v}_2 + O(t^3) \\ \bar{v}_2 + O(t^2) \end{bmatrix}. \tag{2.7}$$

Moreover, if the left singular vector of $\bar{s}_2$ is denoted by $\bar{u}_2$, then the corresponding left singular vector of X_t satisfies

$$u_2^{(t)} = \bar{u}_2 + O(t^2). \tag{2.8}$$

Proof. We shall use the fact that the right singular vectors of X_t are the eigenvectors of

$$A_t \equiv X_t^T X_t = \begin{bmatrix} A_{11} & tA_{21}^T \\ tA_{21} & t^2 A_{22} \end{bmatrix},$$

and the singular values are the nonnegative square roots of the corresponding eigenvalues. The matrix A_0 has two invariant subspaces corresponding to its partitioning. Specifically the columns of

$$\begin{bmatrix} I \\ 0 \end{bmatrix}$$

span an invariant subspace whose eigenvalues are those of A_{11}, while the columns of

$$\begin{bmatrix} 0 \\ I \end{bmatrix}$$

span an invariant subspace whose eigenvalues are zero. The idea of the proof is to use the perturbation theory in [2] to derive expressions for the corresponding invariant subspaces of A_t. Expressions for the individual singular values and singular vectors may then be obtained from the expressions for the invariant subspaces.

Since X has full column rank, A_{11} is nonsingular. It follows [2] that for all sufficiently small t there is a matrix P satisfying

$$(-P\ I)A_t \begin{bmatrix} I \\ P \end{bmatrix} = 0 \tag{2.9}$$

and

$$P = O(t)$$

such that

$$\begin{bmatrix} I \\ P \end{bmatrix}$$

spans an invariant subspace of A_t. If (2.9) is expanded and terms of order greater than t are dropped, the result is

$$P = tA_{21}A_{11}^{-1} + O(t^3) = tB^T + O(t^3). \tag{2.10}$$

The eigenvectors of A_t in the invariant subspace have the form

$$\begin{bmatrix} I \\ P \end{bmatrix} v, \tag{2.11}$$

where v is an eigenvector of

$$(I + P^T P)^{-1/2} (I\ P^T) A \begin{bmatrix} I \\ P \end{bmatrix} (I + P^T P)^{-1/2} = A_{11} + O(t^2). \tag{2.12}$$

If v_1 is a singular vector of X_1 with singular value s_1, then it is an eigenvector of A_{11} with eigenvalue s_1^2. The order t^2 perturbation in (2.12) perturbs s_1^2 by terms of order t^2, from which (2.3) follows. If s_1 is simple, the order t^2 perturbation in (2.12) induces an order t^2 perturbation in v_1. Thus we may take $v = v_1 + O(t^2)$ in (2.11), and (2.4) follows from (2.10). Finally (2.5) follows from the relation $u_1^{(t)} = s_1^{(t)-1} X_t v_1^{(t)}$.

In order to establish (2.6), (2.7), and (2.8), note that the orthogonal complement of the space spanned by the columns of $(I \;\; P^T)^T$ is the complementary invariant subspace of A_t. This subspace is spanned by

$$\begin{bmatrix} -P^T \\ I \end{bmatrix} = \begin{bmatrix} -tB + O(t^3) \\ I \end{bmatrix}.$$

As above, the eigenvectors and eigenvalues of the invariant subspace are to be found from the matrix

$$\begin{aligned} &(I + PP^T)^{-1/2}(-P\ I)A\begin{bmatrix} -P^T \\ I \end{bmatrix}(I + PP^T)^{-1/2} \\ &= t^2\left(-B^T A_{11} B - B^T A_{21}^T - A_{12} B + A_{22}\right) + O(t^4) \\ &= t^2 (X_2 - X_1 B)^T (X_2 - X_1 B) + O(t^4) \\ &= t^2 \overline{X}_2^T \overline{X}_2 + O(t^4). \end{aligned} \tag{2.13}$$

The results now follow from (2.13) by reasoning as above.

Theorem 2.1 divides the singular values of X_t into two classes. The first consists of the singular values of X_1 perturbed by terms of $O(t^2)$. The initial components of the corresponding right singular vector approach the right singular vector v_1 of X_1, while the last components approach zero linearly with t along the direction $B^T v_1$. The left singular vector is the left singular vector of X_1 up to terms of order t^2.

Singular values of the second class approach zero linearly with t. However, they are to be sought in the matrix $t\overline{X}_2$, not tX_2. From (2.1) and (2.2), it follows that

$$\overline{X}_2 = \left[I - X_1 (X_1^T X_1)^{-1} X_1^T\right] X_2.$$

From this it is seen that $\overline{X}_2$ is the projection of X_2 onto the orthogonal complement of the column space of X_1. Singular vectors of the second class behave like those of the first class, except that it is the first components of the right singular vectors that approach zero.

A particularly satisfying feature of these expansions is that the error in any expression is $O(t^2)$ times the order of the expression itself. This suggests that one can expect the asymptotic behavior to set in very quickly.

3. The QR Decomposition. In this section we shall derive expressions for the asymptotic form of the QR decomposition of the matrix X_t defined by (1.5). The results are summarized in the following theorem.

THEOREM 3.1. *Let X be of full column rank and be partitioned as in* (1.4), *where X_{11} is square. Let*

$$R_{11}^T R_{11} = X_{11}^T X_{11}$$

be the Cholesky factorization of $X_{11}^T X_{11}$. *Let*

$$\begin{bmatrix} X_{11} & X_{12} \\ tX_{21} & tX_{22} \end{bmatrix} = \begin{bmatrix} Q_{11}^{(t)} & Q_{21}^{(t)} \\ Q_{12}^{(t)} & Q_{22}^{(t)} \end{bmatrix} \begin{bmatrix} R_{11}^{(t)} & R_{12}^{(t)} \\ 0 & R_{22}^{(t)} \end{bmatrix}$$

be the QR factorization of X_t. *Let*

$$\overline{X}_{22} = X_{22} - X_{21} X_{11}^{-1} X_{12}, \tag{3.1}$$

and let

$$\overline{X}_{22} = \overline{Q}_{22} \overline{R}_{22}$$

be the QR factorization of $\overline{X}_{22}$. *Then*

$$\begin{bmatrix} R_{11}^{(t)} & R_{12}^{(t)} \\ 0 & R_{22}^{(t)} \end{bmatrix} = \begin{bmatrix} R_{11} + O(t^2) & R_{11}^{-1} X_{12} + O(t^2) \\ 0 & t\overline{R}_{22} + O(t^3) \end{bmatrix}. \tag{3.2}$$

Moreover,

$$\begin{bmatrix} Q_{11}^{(t)} \\ Q_{21}^{(t)} \end{bmatrix} = \begin{bmatrix} X_{11} R_{11}^{-1} + O(t^2) \\ tX_{21} R_{11}^{-1} + O(t^3) \end{bmatrix}, \tag{3.3}$$

and

$$\begin{bmatrix} Q_{12}^{(t)} \\ Q_{22}^{(t)} \end{bmatrix} = \begin{bmatrix} -tX_{11}^{-T} X_{21}^T + O(t^3) \\ I + O(t^2) \end{bmatrix} \overline{Q}_{22}. \tag{3.4}$$

Proof. We use the fact that the R-factor of the QR factorization of a matrix X is the Cholesky factor of $X^T X$. In particular,

$$R_{11}^{(t)T} R_{11}^{(t)} = \begin{bmatrix} X_{11} \\ tX_{21} \end{bmatrix}^T \begin{bmatrix} X_{11} \\ tX_{21} \end{bmatrix} = X_{11}^T X_{11} + O(t^2). \tag{3.5}$$

Since the Cholesky decomposition of a nonsingular matrix is a differentiable function of its elements, it follows from (3.5) that $R_{11}^{(t)} = R_{11} + O(t^2)$. Since $R_{12}^{(t)} = R_{11}^{(t)-1} X_{12}$ we have $R_{12}^{(t)} = R_{11}^{-1} X_{12} + O(t^2)$. This takes care of the first row of (3.2). We will return to the expression for $R_{22}^{(t)}$ later. The expression (3.3) is derived from the relation

$$\begin{bmatrix} Q_{11}^{(t)} \\ Q_{21}^{(t)} \end{bmatrix} = \begin{bmatrix} X_{11} \\ tX_{21} \end{bmatrix} R_{11}^{(t)-1}.$$

In order to derive (3.4) and an expression for $R_{22}^{(t)}$, we use the fact that

$$\begin{bmatrix} Q_{12}^{(t)} \\ Q_{22}^{(t)} \end{bmatrix} R_{22}^{(t)}$$

is the projection of $(X_{12}^T \; tX_{22}^T)^T$ onto the orthogonal complement of the space spanned by $(X_{11}^T \; tX_{21}^T)^T$. The projection matrix has the form $I - P$, where

$$\begin{aligned} P &= \begin{bmatrix} X_{11} \\ tX_{21} \end{bmatrix} \left(X_{11}^T X_{11} + t^2 X_{21}^T X_{21} \right)^{-1} \left(X_{11}^T \; tX_{21}^T \right) \\ &= \begin{bmatrix} I \\ tC \end{bmatrix} (I + t^2 C^T C)^{-1} (I \; tC^T), \end{aligned}$$

where we have written $C = X_{21}X_{11}^{-1}$. After some manipulation it follows that

$$(3.6)\qquad (I-P)\begin{bmatrix} X_{12} \\ tX_{22} \end{bmatrix} = \begin{bmatrix} -t^2C^T\overline{X}_{22} + O(t^4) \\ t\overline{X}_{22} + O(t^3) \end{bmatrix}.$$

Thus $R_{22}^{(t)T}R_{22}^{(t)} = t^2\overline{X}_{22}^T\overline{X}_{22} + O(t^4)$, from which it follows that $R_{22}^{(t)} = t\overline{R}_{22} + O(t^3)$. Equation (3.4) follows from (3.6) and the fact that

$$\begin{bmatrix} Q_{12}^{(t)} \\ Q_{22}^{(t)} \end{bmatrix} = (I-P)\begin{bmatrix} X_{12} \\ tX_{22} \end{bmatrix} R_{22}^{(t)-1}.$$

Theorem 3.1 has some analogies with Theorem 2.1. In both theorems the decomposition is divided into two parts, one having a nonzero limit, and the other having a zero limit. In both theorems the first part is obtained from the constant part of the original matrix [e.g., $(X_{11}X_{12})$], while the vanishing part is obtained from a modification of the vanishing part of the original decomposition (e.g., from $\overline{X}_{22}$). In Theorem 2.1 the row space of X is divided into two orthogonal subspaces which become uncoupled in the natural coordinate system of the partition [see (2.4) and (2.7)]. In Theorem 3.1 the same thing happens to the column space of X.

The matrix $\overline{X}_{22}$ defined by (3.1) is a generalization of the Schur complement of X_{11} in X. For square matrices, the Schur complement is what is left over after Gaussian elimination has been performed on X_{11}. Its appearance here can be related to the solution by elimination of the constrained least squares problem

$$(3.7)\qquad \text{minimize}\left\| y - (X_{21}\ X_{22})\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right\|$$

$$(3.8)\qquad \text{subject to } (X_{11}\ X_{12})\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = c.$$

If (3.8) is solved for b_1 and the result is substituted into (3.7), the result is the problem

$$(3.9)\qquad \underset{b_2}{\text{minimize}}\ \left\|(y - X_{21}X_{11}^{-1}c) - \overline{X}_{22}b_2\right\|.$$

On the other hand, the solution may be approximated as in [1] by solving

$$(3.10)\qquad \underset{b_t}{\text{minimize}}\ \left\|\begin{bmatrix} c \\ ty \end{bmatrix} - X_tb_t\right\|$$

for small t. From (3.2) and (3.4) the solution is seen to be

$$(3.11)\qquad b_2^{(t)} = \overline{R}_{22}^{-1}\overline{Q}_{22}^T(y - X_{21}X_{11}^{-1}c) + O(t^2).$$

Since $\overline{R}_{22}^{-1}\overline{Q}_{22}^T$ is the pseudo-inverse of $\overline{X}_{22}$, a comparison of (3.9) and (3.10) shows that $b_2^{(t)}$ differs from b_2 by terms of $O(t^2)$. Thus the solution by weighting is seen as an approximation to the solution by elimination.

This relation throws light on an interesting piece of folklore: namely, that pivoting on column size must be used when solving weighted least squares problems of the form (3.10). The dicta is usually justified by an appeal to the numerical properties of the particular algorithm used to compute the QR decomposition. An alternative line of reasoning goes as follows. The method of elimination will not work if X_{11} is singular, and it can be expected to produce inaccurate results when X_{11} is nearly singular. Since the method of weights mimics the method of elimination, the columns X_t should be interchanged to make X_{11} well-conditioned. Pivoting on column size during the computations of the QR decomposition is an adapative algorithm for doing just this.

Department of Computer Science
University of Maryland
College Park, Maryland 20742

1. C. L. LAWSON & R. J. HANSON, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N. J., 1974.

2. G. W. STEWART, "Error and perturbation bounds for subspaces associated with certain eigenvalue problems," *SIAM Rev.*, v. 15, 1973, pp. 727–767.

3. G. W. STEWART, "Rank degeneracy," *SIAM J. Sci. Statist. Comput.*, v. 5, 1984, pp. 403–413.

4. G. STRANG, *Linear Algebra and Its Applications*, Academic Press, New York, 1976.