University of Maryland College Park

Institute for Advanced Computer Studies TR–92–108
Department of Computer Science TR–2972

Determining Rank in the Presence of Error*

G. W. Stewart†

October, 1992

ABSTRACT

The problem of determining rank in the presence of error occurs in a number of applications. The usual approach is to compute a rank-revealing decomposition and make a decision about the rank by examining the small elements of the decomposition. In this paper we look at three commonly use decompositions: the singular value decomposition, the pivoted QR decomposition, and the URV decomposition.

*This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports`. The report will appear in the proceedings of the NATO Workshop on Large Scale Linear Algebra, Leuven, Belgium, 1992.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742.

Determining Rank in the Presence of Error*

G. W. Stewart†

ABSTRACT

The problem of determining rank in the presence of error occurs in a number of applications. The usual approach is to compute a rank-revealing decomposition and make a decision about the rank by examining the small elements of the decomposition. In this paper we look at three commonly use decompositions: the singular value decomposition, the pivoted QR decomposition, and the URV decomposition.

Introduction

The problem of determining the rank of a matrix has any number of mathematical solutions. For example, if X is an $n \times p$ $(n \geq p)$ of rank k, then X can be reduced by elementary transformations to a row echelon form in which the first k rows of X are linearly independent and the remaining rows are zero. This factorization is perhaps the most widely known example of a *rank-revealing decomposition*—a decomposition in which the rank can be read off from the pattern of zero and nonzero elements. There are, of course, many other rank revealing decompositions; e.g., the singular value decomposition, the QRP decomposition, and a variety of complete orthogonal factorizations.

The problem is far more difficult when the elements of X are contaminated with error, so that instead of X we observe

$$\tilde{X} = X + E,$$

where E is unknown. In this case we must determine the rank k of the original matrix from the contaminated matrix $\tilde{X}$. The usual approach is to compute a rank-revealing decomposition of $\tilde{X}$. Since, in general, $\tilde{X}$ will be of full rank, the decomposition will not reveal the rank by the structure of its zero elements. Instead one looks at the structure of the "small" elements in the hope that they will say something about the rank of X. However, there are several difficulties with this general approach, which we now list briefly.

*This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports`. The report will appear in the proceedings of the NATO Workshop on Large Scale Linear Algebra, Leuven, Belgium, 1992.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742.

1. If we compute a decomposition $\tilde{Z} = \tilde{U}^{\mathrm{T}}\tilde{X}\tilde{V}$ of $\tilde{X}$ corresponding to the rank-revealing decomposition $Z = U^{\mathrm{T}}XV$ of X, there is no guarantee that $\tilde{Z}$ have small elements in place of the revealing zeros of Z. As we shall see, this is a problem even when the transformations U and V are orthogonal.

2. We must have some knowledge of E to say when elements in a rank-revealing decomposition are "small". Usually this knowledge is in the form of an estimate of a norm of E, or the size of a "typical" element of E, or even a statistical distribution of the elements of E. Whatever form this knowledge takes, it must come from outside sources, i.e. the nature of the application.

3. We also need to know something about X. For example, if $\tilde{X} = \mathrm{diag}(1, 10^{-3})$ and we know that $\|E\| \cong 10^2$ in the spectral norm, we cannot say that X has rank one — only that X being of rank one is not inconsistent with what we know. In order to make a stronger statement we need to know, say, that the smallest nonzero singular value of X is greater than 10^{-2}.

4. The results of a rank determination can vary with the scaling. For example, if the last row of the matrix X in the preceding item of this list is multiplied by 10^3, then X becomes the identity, and the most we can say about E is that $\|E\| \cong 10$. In this case the data are consistent with X being the zero matrix! The usual fix is to attempt to scale $\tilde{X}$ so that the elements of E are roughly the same size. But this is not always easily done.

5. In most applications, rank determination is only a beginning. What is done subsequently often requires a knowledge of the column or null spaces of X or X^{T}. Thus any rank-revealing decomposition must produce approximations to these spaces.

6. In many applications the rows of X are not fixed but change as rows are added and deleted. Since rank-revealing decompositions are usually too expensive to recompute *ab initio*, our decomposition should be updatable.

Keeping these difficulties in mind, we are going discuss three decompositions that are used to determine the rank of a matrix: the singular value decomposition, the pivoted QR decomposition (also called the QRP decomposition), and an intermediary called the URV decomposition. All three are based on orthogonal transformations, which have the desirable property that they cannot magnify the error in X. In the next three sections we will describe these decompositions and discuss their numerical properties. In the concluding section we will treat the problem of using these decompositions to determine rank in the presence of error.

1. The Singular Value Decomposition

The singular value decomposition is the *crème de la crème* among rank-revealing decomposition. It has the form

$$U^{\mathrm{T}} X V = \begin{pmatrix} \Sigma \\ 0 \end{pmatrix}, \tag{1}$$

where U and V are orthogonal and

$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p),$$

with

$$\sigma_1 \geq \cdots \geq \sigma_k > 0 = \sigma_{k+1} = \cdots = \sigma_p.$$

Thus the rank k of X is revealed by the fact that its $p-k$ largest singular values are nonzero while its k smallest singular values are zero.

The singular value decomposition easily provides orthonormal bases for range and null spaces associated with X. Specifically, if we partition

$$U = (U_1\ U_2) \quad \text{and} \quad V = (V_1\ V_2), \tag{2}$$

then:

1. The columns of U_1 form an orthonormal basis for the column space of X.
2. The columns of U_2 form an orthonormal basis for the null space of X^{T}.
3. The columns of V_1 form an orthonormal basis for the column space of X^{T}.
4. The columns of V_2 form an orthonormal basis for the null space of X.

The singular value decomposition behaves well in the presence of error. Specifically, if

$$\tilde{U} \tilde{X} \tilde{V}^{\mathrm{T}} = \begin{pmatrix} \tilde{\Sigma} \\ 0 \end{pmatrix}$$

is the singular value decomposition of $\tilde{X}$, then it follows from Schmidt's theorem that

$$\tilde{\sigma}_{k+1}^2 + \cdots + \tilde{\sigma}_p^2 \leq \|E\|_{\mathrm{F}}^2, \tag{3}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. Thus the singular value decomposition of $\tilde{X}$ reveals the rank in the sense that the sum of squares of its $p-k$ smallest singular values are bounded by the Frobenius norm of E.[1] Moreover, the spaces spanned by

[1] Actually, each of the $n-k$ smallest singular values are bounded by the spectral norm of E; however, this result is less useful in practice than (3).

$\tilde{U}_1$, $\tilde{U}_2$, $\tilde{V}_1$, and $\tilde{V}_2$ are approximations to the subspaces listed above that are accurate to about $\sigma_k^{-1}\|E\|$. Thus if σ_k is reasonably large compared to E—i.e., the problem has a favorable signal to noise ratio—the singular value decomposition provides good approximations to the desired subspaces.

The singular value decomposition can be computed in many ways, among which the following three are the most common.

1. Reduce X to bidiagonal form by two sided orthogonal transformations and reduce the bidiagonal form to diagonal form by a variant of the QR algorithm.

2. Reduce X to triangular form by transformations applied on the right (i.e., compute the QR decomposition of X). Then compute the singular value decomposition of the triangular matrix.

3. Compute the eigendecomposition of the symmetric cross-product matrix $X^{\mathrm{T}}X$.

The first method is by far the most expensive, though it is standard for one-shot jobs. The other two approaches have the advantage that the intermediate decompositions can be updated in $O(p^2)$ time, the third trivially and very quickly. It is true that when we update in this way, we loose the ability to compute the matrix U; but in most updating applications U is not required. It is sometimes objected that computing the singular value decomposition via the cross-product matrix is numerically unstable. But these instabilities only become important when σ_k^2/σ_1^2 approaches the rounding unit of the computer arithmetic—something that seldom happens in the presence of errors other than rounding error.

In the updating game, no matter which of the above algorithms is used, one is left with the problem of updating a singular value decomposition or eigendecomposition of a square matrix of order p. Unfortunately, no algorithms that perform these updates in less that $O(p^3)$ time are known, a fact that severly restricts the use of the singular value decomposition in real-time applications. Recent work has focused on maintaining an approximate diagonal form that is good enough for practical purposes. However, an alternative to which we now turn, is to work with more computationally tractable decompositions.

2. QRP Decompositions

The pivoted QR decomposition, or the QRP decomposition as it will be called here, is actually a class of decompositions. Specifically, there is a permutation matrix P and an orthogonal matrix Q such that

$$Q^{\mathrm{T}}XP = \begin{pmatrix} R \\ 0 \end{pmatrix}, \tag{4}$$

where

$$R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & 0 \end{pmatrix}$$

with R_{11} an upper triangular matrix of order k having positive diagonal elements. The decomposition is not unique, since P can be any permutation matrix (also called a *pivot matrix*) such that the first k columns of XP are unique. Once P has been determined, however, the matrices R_{11} and R_{12} are uniquely determined, as are the first k columns of Q.

If we partition $XP = (X_1\ X_2)$ and $Q = (Q_1\ Q_2)$, where X_1 and Q_1 have k columns, then

1. The columns of X_1 and Q_1 form an orthonormal basis for the column space of X.
2. The columns of Q_2 form an orthonormal basis for the null space of X^{T}.
3. The columns of $(R_{11}\ R_{12})^{\mathrm{T}}$ form a (nonorthonormal) basis for the column space of X^{T}.
4. The columns of $(-R_{12}^{\mathrm{T}} R_{11}^{-\mathrm{T}}\ I)^{\mathrm{T}}$ form a (nonorthonormal) basis for the null space of X.

This list illustrates some of the strengths and weaknesses of QRP decompositions. As far as the row space of X and the null space of X^{T} are concerned, the matrix Q is entirely analogous to the matrix U of the singular value decompositions. Moreover, the first k columns of XP form a basis for the columns space of X; i.e., the QRP decomposition picks out a set of linearly independent columns, unlike the singular value decomposition, which merely furnishes a basis for the column space. Unfortunately, there is no analogue of the matrix V, and the corresponding bases, which must be obtained from R, are not orthonormal. Moreover, the basis for the null space of X requires additional computation for its formation. This is particularly unfortunate, since many applications require an orthonormal basis for this subspace.

In the presence of error, we should like to determine a permutation matrix $\tilde{P}$ and an orthogonal matrix $\tilde{Q}$ such that

$$\tilde{Q}^{\mathrm{T}} \tilde{X} \tilde{P} = \begin{pmatrix} \tilde{R}_{11} & \tilde{R}_{12} \\ 0 & G_{22} \end{pmatrix}, \tag{5}$$

where R_{11} is upper triangular of order k with positive diagonal elements, and G_{22} is a triangular matrix satisfying

$$\|G_{22}\| = O(\|E\|).$$

Note that this amounts to finding a suitable pivot matrix P, since once P is chosen, the rest of the decomposition is essentially unique.

The question of the existence of a rank-revealing QRP decomposition in the sense of the preceding paragraph has only recently been answered in the affirmative, Unfortunately, the proof is not constructive, and the problem of efficiently computing a provably rank-revealing QRP decomposition is still an active area of research (for more see the notes and references at the end of the paper).

For practical purposes, however, almost any sensible strategy will work. The standard algorithm is unitary triangularization with pivoting on the column of largest norm. This procedure is mathematically (though not numerically) equivalent to the Gram-Schmidt algorithm in which the largest projected vector is the next to enter the orthogonalization.

An alternative is the rank-revealing algorithm of T. Chan, which starts from an unpivoted QR decomposition and moves linearly dependent columns of R to the end. The first step of this algorithm is typical. A condition estimator is used to find a vector v of norm one such that $\epsilon = \|Rv\|$ is small. A permutation matrix $\hat{P}$ and an orthogonal matrix $\hat{Q}$ are determined so that

1. the last component of $\hat{v} = \hat{P}^{\mathrm{T}}v$ is the largest,
2. $\hat{R} = \hat{Q}^{\mathrm{T}}R\hat{P}$ is upper triangular.

Since the magnitude of the last component of $\hat{v}$ is not less than $1/\sqrt{p}$ and $\|\hat{R}\hat{v}\| = \epsilon$, it follows that $|\hat{r}_{pp}| \leq \sqrt{p}\epsilon$; i.e., $\hat{R}$ reveals the degeneracy in the rank of R. Unfortunately, when this procedure is iterated, the provable bound on the size of the resulting G_{22} grows exponentially, though in practice the algorithms works well enough.

At present there seems to be no efficient algorithm for updating a rank-revealing QRP algorithm.

3. URV and ULV Decompositions

Although the simplicity of QRP decompositions makes them attractive, the fact that P must be a permutation matrix is a combinatorial constraint that makes analysis difficult and algorithms hard to come by. In this section we will consider another class of decompositions that mitigates these problems by relaxing the restriction on P.

The decompositions are based on what is sometimes called a complete orthogonal decomposition. If X is exactly of rank k, there are orthogonal matrices U and V such that

$$U^{\mathrm{T}}XV = \begin{pmatrix} R_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

where R is an upper triangular matrix of order k having positive diagonal elements. Such a decomposition is not unique: the singular value decomposition is an extreme example. However, by relaxing the restriction that the decomposition be diagonal, we introduce extra degrees of freedom in U and V that make for flexability.

This flexability does not imply a loss of information. If we partition U and V as in (2), then the statements following that equation remain true. In other words, the decomposition provides orthonormal bases for the range and null spaces associated with X.

A rank-revealing URV decomposition of $\tilde{X}$ is a decomposition of the form

$$\tilde{U}^{\mathrm{T}}\tilde{X}\tilde{V} = \begin{pmatrix} R_{11} & F_{12} \\ 0 & G_{22} \end{pmatrix},$$

where, as usual, R is an upper triangular matrix of order k having positive diagonal elements and F_{12} and G_{22} are of order $\|E\|$. Unlike the QRP decomposition, the URV decomposition can be made provably rank revealing. The process begins like Chan's method by reducing X to a triangular form R and using a condition estimator to find a vector v of norm one such that $\epsilon = \|Rv\|$ is small. Orthogonal matrices $\hat{U}$ and $\hat{V}$ are determined so that

1. $\hat{V}^{\mathrm{T}}v = \mathbf{e}_p$, where $\mathbf{e}_p$ is the vector whose last component is one and whose other components are zero,

2. $\hat{R} = \hat{U}^{\mathrm{T}}R\hat{V}$ is upper triangular.

It then follows that $\hat{R}$ has the form

$$\hat{R} = \begin{pmatrix} \hat{R}_{11} & f_{12} \\ 0 & \gamma_{22} \end{pmatrix},$$

where $\|(f_{12}^{\mathrm{T}}\ \gamma_{22})\| = \epsilon$. Thus $\hat{R}$ reveals the rank degeneracy in R. However, instead of just the (p,p)-element of $\hat{R}$ being small, as in the QRP decomposition, the entire last column of $\hat{R}$ is small. This fact allows us to iterate the process on $\hat{R}_{11}$ to get a provably rank-revealing decomposition of X. At each stage, a block variant of the QR algorithm, can be applied to further reduce the size of f_{12}.

An attractive feature of rank-revealing URV decompositions is that they can be updated in $O(p^2)$ time. Moreover, the updating algorithm can be implemented in $O(p)$ time on a linear array of p processors. The updating algorithm can be started with the zero matrix, so that there is no need to compute an initial decomposition.

There is also a rank-revealing ULV decomposition, in which the target matrix is lower triangular. Surprisingly, these decompositions are not mere variants of one another but have different mathematical algorithmic properties — an area for future research.

4. Rank Determination

The term "rank-revealing decomposition" is something of a misnomer, since it implies that the decomposition automatically reveals rank. As we indicated in Items 2 and 3 of the list in the introduction, a decomposition alone is never sufficient: we need to know something about the error, and perhaps also about the original matrix. In this section, we will discuss the how to use our three rank-revealing decompositions to determine rank in the presence of errors.

4.1. The Singular Value Decomposition

Suppose for the moment that we know the rank k of X and desire to estimate the matrix X from $\tilde{X}$. A natural procedure is to try to approximate $\tilde{X}$ by a matrix of rank k in the least squares sense. Otherwise put, our estimate of X is a matrix $\hat{X}$ that satisfies

$$\|\tilde{X} - \hat{X}\|_{\mathrm{F}} = \min_{\mathrm{rank}(Y)\leq k} \|\tilde{X} - Y\|_{\mathrm{F}}.$$

Fischer's theorem says that $\hat{X}$ exists and that

$$\|\tilde{X} - \hat{X}\|_{\mathrm{F}}^2 = \tilde{\sigma}_{k+1}^2 + \cdots + \tilde{\sigma}_n^2 \stackrel{\mathrm{def}}{=} \tilde{\tau}_k^2 \tag{6}$$

For our purposes the most important consequence of Fischer's theorem is the following. Since $\hat{X}$ is minimizing, the right hand side of (6) can only increase when we replace $\hat{X}$ by X. Consequently,

$$\|E\|_{\mathrm{F}}^2 \geq \|X - \hat{X}\|_{\mathrm{F}}^2 = \tilde{\tau}_k^2.$$

The implication is that if $\|E\|_{\mathrm{F}}$ is smaller than $\tilde{\tau}_k^2$ then X could not possibly been of rank k. Thus a natural choice of k is the smallest integer such that

$$\|E\|_{\mathrm{F}}^2 \geq \tilde{\tau}_k^2.$$

This strategy works well, provided the errors are well scaled and σ_k is well above the error level. In this case $\tilde{\tau}_k^2$ remains below $\|E\|_{\mathrm{F}}^2$, but the presence of $\tilde{\sigma}_k$ forces the sum $\tilde{\tau}_{k-1}^2$ to be larger than $\|E\|_{\mathrm{F}}^2$ (recall that from the perturbation theory for singular values, $\tilde{\sigma}_k \geq \sigma_k - \|E\|$, so that $\tilde{\sigma}_k$ is large along with σ_k).

If we are willing to assume more about E, we can refine our procedure. Let us suppose that the elements of E are uncorrelated random variables with mean zero and standard deviation ϵ. Because the matrices U and V in the singular value decomposition (1) of X are orthogonal, the elements of the matrices H_{ij} in

$$U^{\mathrm{T}}\tilde{X}V = U^{\mathrm{T}}XV + U^{\mathrm{T}}EV \equiv \begin{pmatrix} \Sigma_1 + H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \tag{7}$$

are also uncorrelated with mean zero and standard deviation ϵ. Now if σ_k is large compared with $\|E\|$, then

$$\tilde{\tau}_k^2 = \|H_{22}\|_{\mathrm{F}}^2 + O(\|E\|^3). \tag{8}$$

It follows that that the average value of $\tilde{\sigma}_{k+1}^2 + \cdots + \tilde{\sigma}_p^2$ is approximated by

$$\mathbf{E}(\tilde{\tau}_k^2) \cong (n-k)(p-k)\epsilon^2.$$

Consequently, we should choose k to be the smallest integer such that

$$\tilde{\tau}_k^2 < \phi \cdot (n-k)(p-k)\epsilon^2. \tag{9}$$

The number ϕ in (9) is a fudge factor that compensates for the fact that $\tilde{\tau}_k^2$ will often be larger than its mean. If it is too small, i.e. too near one, the test will tend to overestimate the rank. If it is too large, the test will tend to underestimate the rank. If we know the distribution of the elements of E and the value of σ_k, we can choose ϕ to trade these errors off against one another. However, it seldom happens in practice that we have such precise information, and the value of ϕ must usually be chosen on the basis of experience.

The statistical assumptions about the elements of E—that they are uncorrelated with mean zero and common standard deviation ϵ—correspond to the equal error scaling mentioned in Item 4 of the list in the introduction. If these assumptions are not satisfied, it may be possible to scale the problem so that they are. For example, if the rows of E are uncorrelated with mean zero and dispersion (variance) matrix Δ, then the elements of $E\Delta^{-\frac{1}{2}}$ are uncorrelated with mean zero and standard deviation one. This process is sometimes called "whitening" the noise. However, it cannot be applied when Δ is singular (e.g., when a column of $\tilde{X}$ is without error). What to do in such situations is imperfectly understood.

Finally, it is important not to lay too much stress on detailed statistical assumptions about the error. Informally all that is needed is for the elements of E to be roughly the same size ϵ and to remain so under unrelated orthogonal transformations. In that case the elements of the matrix $H_{22} = U_2^{\mathrm{T}} E V_2$ will be of roughly of size ϵ and $\|H_{22}\|_{\mathrm{F}}^2$ will be approximately $(n-k)(p-k)\epsilon^2$. This is all that is required for the validity of the test (9).

4.2. QRP Decompositions

Rank determination with QRP decompositions is not as straightforward as it is with the singular value decomposition. In the first place there is no analogue of Fisher's theorem for the decomposition. Moreover, if we attempt to repeat the development that produced the test (9) we run into difficulties.

To see why, let us apply the transformations Q and P to the matrix $\tilde{X}$. The result is

$$U^{\mathrm{T}}\tilde{X}P = \begin{pmatrix} R_{11}+H_{11} & R_{12}+H_{12} \\ H_{21} & H_{22} \end{pmatrix}. \tag{10}$$

Now the matrix H_{22} is quite tractable. Under our assumptions about the distribution of the elements of E, the expectation of $\|H_{22}\|_{\mathrm{F}}^2$ is $(n-k)(p-k)\epsilon^2$. Unfortunately, $\|H_{22}\|_{\mathrm{F}}$ is not an approximation of $\|G_{22}\|_{\mathrm{F}}$, which is what we have to work with. The reason is that $U^{\mathrm{T}}\tilde{X}P$ is not in triangular form, and when we reduce it to triangular form, the matrix G_{22} [c.f. (5)] becomes contaminated by the elements of the large matrix R_{12}. (The same sort of thing does not occur with the singular value decomposition because *both* off-diagonal blocks in (7) are small.)

Fortunately, we can still approximate the expected value of $\|G_{22}\|_{\mathrm{F}}$. Specifically, the norm of G_{22} is the norm last $n-k$ rows of the projection of the last column of (10) onto the orthogonal complement of the space spanned by the first column. Explicitly, the norm of G_{22} is the norm of the matrix

$$H_{22} - H_{21}[(R_{11}+H_{11})^{\mathrm{T}}(R_{11}+H_{11}) + H_{21}^{\mathrm{T}}H_{21}]^{-1}[(R_{11}+H_{11})^{\mathrm{T}}(R_{12}+H_{12}) + H_{21}^{\mathrm{T}}H_{22}].$$

If we ignore higher order terms, we get

$$\|G_{22}\|_{\mathrm{F}}^2 \cong \|H_{22} - H_{21}R_{11}^{-1}R_{12}\|_{\mathrm{F}}^2. \tag{11}$$

Consequently, the expected value of $\|G_{22}\|_{\mathrm{F}}^2$ should be approximately

$$(n-k)\mathrm{trace}[I_{p-k} + R_{12}^{\mathrm{T}}(R_{11}^{\mathrm{T}}R_{11})^{-1}R_{12}]\epsilon^2. \tag{12}$$

Of course we do not know R_{11} and R_{12}; however, the computed matrices $\tilde{R}_{11}$ and $\tilde{R}_{12}$ are small perturbations of the originals and can be used in their place. Thus, our test is to choose k to be the smallest integer such that

$$\|G_{22}\|_{\mathrm{F}}^2 < \phi\cdot(n-k)\mathrm{trace}[I_{p-k} + \tilde{R}_{12}^{\mathrm{T}}(\tilde{R}_{11}^{\mathrm{T}}\tilde{R}_{11})^{-1}\tilde{R}_{12}]\epsilon^2, \tag{13}$$

where ϕ is the usual fudge factor.

4.3. URV Decompositions

The same argument that was used to show (8) can be used to show that for the URV decomposition

$$\tilde{\tau}_k^2 = \|G_{22}\|_{\mathrm{F}}^2 + O(\|E\|^3).$$

Consequently an appropriate test for the URV decomposition is to choose the smallest k such that

$$\|G_{22}\|_{\mathrm{F}}^2 < \phi\cdot(n-k)(p-k)\epsilon^2.$$

5. Notes and References

The problem of rank determination in the presence of error arises in a number of applications: e.g., variable selection in statistics and engineering [7, 47, 61], direction of arrival estimation in signal processing [1, 58, 59], and the projection of ill-conditioned problems onto manifolds where they become well conditioned [51, 21, 22]. In many instances, the original matrix X is not exactly of rank k as we have described it in the introduction. Instead physical approximations or infelicities in the model make X only approximately of rank k, though the deviation must be less than the error for the techniques described here to have approximate validity.

Closely related, but of a different flavor, is the problem of regularizing the ill-posed problems which arise from discretizations of compact or unbounded operators [24, 32, 45, 57, 71, 75].

The fact that something must be known about the errors in order to make statements about rank is a commonplace in areas like signal processing, where the errors are relatively large, or statistics, where there is a vast literature under the heading "errors in the variables" [2, 3, 5, 4, 8, 20, 34, 48, 65], or numerical analysis, where there is a growing literature under the heading "total least squares" [31, 41, 63, 72, 74].

Although the updating of least squares solutions goes back to Gauss [35], the updating of decompositions seems to have arisen in linear program, where the inverse basis matrix must be updated [19]. A closely related problem is that of downdating — the removal of rows from X — a process that is also called windowing. The literature on updating and downdating is too voluminous to survey here.

The singular value decomposition dates to the last half of the nineteenth century (for a history see [67]). The theorem cited here as Schmidt's theorem [60], is often attributed to Eckart and Young [29], who rediscovered it thirty years later. The popularity of the singular value decomposition in numerical analysis is due to Golub and Kahan [38].

Until recently reduction to bidiagonal form followed by a variant of the QR algorithm, due to Golub [40], has been the standard way to compute the decomposition. Recently new algorithms for reducing the bidiagonal matrix have been proposed [25, 30]. The idea of first computing the QR decomposition has been exploited by Chan [12, 11]. Beltrami [6] first established the existence of the singular value decomposition in 1873 by computing the eigendecomposition of the cross-product matrix, and this is still a popular way of doing things in some disciplines. In fact, sometimes the singular value decomposition completely disappears.

Algorithms for updating the singular value decomposition have been given in [10, 18]; however, they require $O(p^3)$ operations, the same as required to compute the decomposition from scratch. Iterative algorithms that maintain an approximate factorization may be found in [54, 55, 56].

Formulas for the discrete version of the Gram-Schmidt algorithm can be found in the first supplement to Laplace's *Théoria Analytique des Probabilities* [52]; however, Laplace was after an expression for the variance of a regression parameter and did not regard his formulas as a computational device. Gram [44] and Schmidt [60] orthogonalized series of functions: Gram by determinantal expressions (hence the Gramian matrix) and Schmidt by the now classic algorithm. The use of orthogonal transformations to compute the decomposition is due to Householder [50], Bogert and Burris [9], and Golub [36]. The last mentioned work also contains the notion of column pivoting and the first updating algorithm for the QR decomposition. The name QR decomposition is from Francis's QR algorithm [33], which uses the decomposition.

Although pivoting for column size while computing the QR decomposition has long been regarded as a reliable way of determining rank (e.g., see [39, 62]), Chan [13] was the first to give bounds for a rank-revealing decomposition (the descriptive phrase "rank revealing" was coined by him). Unfortunately, the bounds were exponential in the defect $p - k$ in the rank. In fact, only recently have Hong and Pan [49] established the existence of a rank revealing QR decomposition. Although their approach is not constructive, Chandrasekaran and Ipsen [14] have given an algorithm, which unfortunately has combinatorial complexity (this paper is an excellent source for other pivoting strategies that have appeared in the literature). In a personal communication and Pan and Tang have described and algorithm that requires less work.

It is important to distinguish the sense in which the theory of Hong and Pan and the algorithms mentioned above are rank revealing. Both take an integer k and produce a permutation that reveals if there is a gap between the kth and $(k+1)$th singular value. Change k and the permutation changes, so that the rank is not necessarily revealed for all k simultaneously.

The ability to cheaply compute an approximate null vector of a triangular matrix — a topic which goes under the slightly misleading name of "conditions estimation" — is fundamental to some algorithms for computing a rank-revealing QRP decomposition as well as the URV and ULV decompositions. Although the first such algorithm is found in [43], it was LINPACK [27] that popularized the idea. For a survey with references see [46].

URV and ULV decompositions [68, 66] had their genesis in the author's unsuccessful attempt to update a rank-revealing QR factorization. A refinement step, which tends to decrease the size of the off-diagonal elements has been analyzed in [53] (see also [56, 15, 28]). A parallel implementation of the updating algorithm is described in [69].

The methods treated here are not the only ones for revealing rank. For example, methods based on the Lanczos algorithm have been proposed for the case where the rank is small [17, 77, 78].

The perturbation of singular values, including Fischer's theorem, is surveyed in [70]. The relation (8) is a consequence of theorems in [53]. The approach to rank determination followed here is rather crude, suitable for the crude models and data one can expect in practice. However, if one can assume normality, then $\tilde{\tau}_k^2$ is approximately χ^2, a fact that can be used to determine a value for the fudge facter ϕ in (9). More generally the singular values $\sigma_{k+1}^2, \ldots \sigma_p^2$ are approximately the eigenvalues of a Wishart matrix, whose distributions are known (e.g., see [26, 16]).

The problem of poorly scaled errors is closely related to the problem of artificial ill-conditioning, which is discussed in [64]. The equal error scaling advocated there is the equivalent of noise whitening. One solution to the problem of constrained errors is to project the problem onto a submanifold where the errors can be whitened [23, 37, 73].

The results on testing QRP decompositions appear to be new. The consequence of (11) and (12) are that $\|G_{22}\|_F^2$ will tend to be larger than $\tilde{\rho}_k^2$. Comparing (9) and (13), we see that the latter has been increased to compensate for this fact.

References

[1] G. Adams, M. F. Griffin, and G. W. Stewart. Direction-of-arrival estimation using the rank-revealing URV decomposition. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Washington, DC, 1991. IEEE.

[2] T. W. Anderson. Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics. *Journal of the Royal Statistical Society*, 38:1–31, 1976.

[3] A. E. Beaton, D. B. Rubin, and J. L. Barone. The acceptability of regression solutions: Another look at computational accuracy. *Journal of the American Statistical Association*, 71:158–168, 1976.

[4] D. A. Belsley. Assessing the presence of harmful collinearity and other forms of weak data through a test for signal-to-noise. *Journal of Econometrics*, 20:211–253, 1982.

[5] D. A. Belsley, A. E. Kuh, and R. E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* John Wiley and Sons, New York, 1980.

[6] E. Beltrami. Sulle funzioni bilineari. *Giornale di Matematiche ad Uso degli Studenti Delle Universita*, 11:98–106, 1873. An English translation by D. Boley is available as University of Minnesota, Department of Computer Science, Technical Report 90–37, 1990.

[7] K. N. Berk. Comparing subset regression procedures. *Technometrics*, 20:1–6, 1978.

[8] J. Berkson. Are there two regressions. *Journal of the American Statistical Association*, 45:164–180, 1950.

[9] D. Bogert and W. R. Burris. Comparison of least squares algorithms. Report ORNL-3499, Neutron Physics Division, Oak Ridge National Laboratory, 1963. Vol. 1, Sec. 5.5.

[10] J. R. Bunch and C. P. Nielsen. Updating the singular value decomposition. *Numerische Mathematik*, 31:111–129, 1978.

[11] T. F. Chan. Algorithm 581: An improved algorithm for computing the singular value decomposition. *ACM Transactions on Mathematical Software*, 8:84–88, 1982.

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

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

[14] S. Chandrasekaran and I. Ipsen. Perturbation theory for the solution of systems of linear equations. Research Report YALEU/DCS/RR-866, Department of Computer Science, Yale University, 1991.

[15] S. Chandrasekaran and I. Ipsen. Analysis of a QR algorithm for computing singular values. Research Report YALEU/DCS/RR-917, Department of Computer Science, Yale University, 1992.

[16] R. Choudary, Hanumara, and W. A. Thompson Jr. Percentage points of the extreme roots of a Wishart matrix. *Biometrika*, 55:505–512, 1968.

[17] P. Comon and G. H. Golub. Tracking a few extreme singular values and vectors in signal processing. *Proc. IEEE*, 78:1327–1343, 1990.

[18] J. J. M. Cuppen. The singular value decomposition in product form. *SIAM Journal on Scientific and Statistical Computing*, 4:216–222, 1983. Cited in [42].

[19] G. B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, New Jersey, 1963.

[20] R. B. Davies and B. Hutton. The effects of errors in the independent variables in linear regression. *Biometrika*, 62:383–391, 1975.

[21] J. Demmel. *A Numerical Analyst's Jordan Canonical Form*. PhD thesis, Department of Computer Science, University of California at Berleley, 1983. Center for Pure and Applied Mathematics Technical Report PAM-156.

[22] J. Demmel. On condition numbers and the distance to the nearest ill-posed problem. *Numerische Mathematik*, 51:251–290, 1987.

[23] J. Demmel. The smallest perturbation of a submatrix which lowers the rank and constrained total least squares problems. *SIAM Journal on Numerical Analysis*, 24:199–206, 1987.

[24] J. Demmel and B. Kågström. Accurate solutions of ill-posed problems in control theory. *SIAM Journal on Matrix Analysis and Applications*, 9:126–145, 1988.

[25] J. Demmel and W. Kahan. Accurate singular values of bidiagonal matrices. *SIAM Journal on Scientific and Statistical Computing*, 11:873–912, 1989.

[26] A. P. Dempster. *Elements of Continuous Multivariate Analysis*. Addison-Wesley, Reading, Massachusetts, 1969.

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

[28] E. M. Dowling, L. P. Ammann, and R. D. DeGroat. A tqr-iteration based adaptive svd for real time angle and frequency tracking. Erik Johnson School of Engineering and Computer Science, The University of Texas at Dallas. Manuscript submitted to IEEE Transactions on Signal Processing, 1992.

[29] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.

[30] K. V. Fernando and B. Parlett. Accurate singular values and differential qd algorithms. Technical Report PAM–554, Department of Mathematics, University of California, Berkeley, 1992.

[31] R. D. Fierro and J. R. Bunch. Multicollinearity and total least squares. Preprint Series 977, Institute for Mathematics and Its Applications, 1992.

[32] H. E. Fleming. Equivalence of regularization and truncated iteration in the solution of ill-posed image reconstruction problems. *Linear Algebra and Its Applicaations*, 130:133–150, 1990.

[33] J. G. F. Francis. The QR transformation, parts I and II. *Computer Journal*, 4:265–271, 332–345, 1961, 1962.

[34] W. A. Fuller. *Measurement Error Models*. John Wiley, New York, 1987.

[35] C. F. Gauss. Theoria combinationis observationum erroribus minimis obnoxiae, pars posterior. In *Werke, IV*, pages 27–53. Königlichen Gesellshaft der Wissenschaften zu Göttingin (1880), 1823.

[36] G. H. Golub. Numerical methods for solving least squares problems. *Numerische Mathematik*, 7:206–216, 1965.

[37] G. H. Golub, A. Hoffman, and G. W. Stewart. A generalization of the Eckart-Young matrix approximation theorem. *Linear Algebra and Its Applications*, 88/89:317–327, 1987.

[38] G. H. Golub and W. Kahan. Calculating the singular values and pseudo-inverse of a matrix. *SIAM Journal on Numerical Analysis*, 2:205–224, 1965.

[39] G. H. Golub, V. Klema, and G. W. Stewart. Rank degeneracy and least squares problems. Technical Report TR-456, Department of Computer Science, University of Maryland, 1976.

[40] G. H. Golub and C. Reinsch. Singular value decomposition and least squares solution. *Numerische Mathematik*, 14:403–420, 1970. Also in [76, pp.134–151].

[41] G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17:883–893, 1980.

[42] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, 2nd edition, 1989.

[43] W. B. Gragg and G. W. Stewart. A stable variant of the secant method for solving nonlinear equations. *SIAM Journal on Numerical Analysis*, 13:880–903, 1976.

[44] J. P. Gram. Über die Entwicklung reeler Functionen in Reihen mittelst der Methode der kleinsten Quadrate. *Journal für die reine und angewandte Mathematik*, 94:41–73, 1883.

[45] P. C. Hansen. Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank. *SIAM Journal on Scientific and Statistical Computing*, 11:503–518, 1990.

[46] N. J. Higham. A survey of condition number estimation for triangular matrices. *SIAM Review*, 29:575–596, 1987.

[47] R. R. Hocking. The analysis and selection of variables in linear regression. *Biometrics*, 32:1–49, 1976.

[48] S. D. Hodges and P. G. Moore. Data uncertainties and least squares regression. *Applied Statistics*, 21:185–195, 1972.

[49] Y. P. Hong and C.-T. Pan. Rank-revealing QR factorizations and the singular value decomposition. *Mathematics of Computation*, 58:213–232, 1992.

[50] A. S. Householder. Unitary triangularization of a nonsymmetric matrix. *Journal of the ACM*, 5:339–342, 1958.

[51] W. Kahan. Conserving confluence curbs ill-conditioning. Technical Report 6, Computer Science Department, University of California, Berkeley, 1972.

[52] P. S. Laplace. *Théoria analytique des probabilities (3rd ed.) premier supplement: Sur l'application du calcul des probabilités a la philosophie naturelle. Oeuvres, v.7.* Gauthier-Villars, 1820. Supplement published before 1820.

[53] R. Mathias and G. W. Stewart. A block qr algorithm and the singular value decomposition. Technical Report CS-TR 2626, Department of Computer Science, University of Maryland, College Park, 1992. To appear in *Linear Algebra adn Its Applications*.

[54] M. Moonen. *Jacobi-Type Updating Algorithms for Signal Processing, Systems Identification and Control.* PhD thesis, Katholieke Universiteit Leuven, 1990.

[55] M. Moonen, P. Van Dooren, and J. Vandewalle. Combined Jacobi-type algorithms in signal processing. In R. J. Vaccaro, editor, *SVD and Signal Processing, II*, pages 177–188, Amsterdam, 1991. Elsevier Science Publishers.

[56] M. Moonen, P. Van Dooren, and F. Vanpoucke. On the QR algorithm and updating the SVD and URV decomposition in parallel. ESAT Katholieke Universiteit Leuven, 1992.

[57] D. P. O'Leary and J. A. Simmons. A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems. *SIAM Journal on Scientific and Statistical Computing*, 2:474–489, 1981.

[58] S. Prasac and B. Chandna. Direction-of-arrival estimation using rank revealing QR factorization. *IEEE Transactions on Signal Processing*, 39:1224–1229, 1991. Citation communicated by Per Christian Hansen.

[59] R. Roy and T. Kailath. ESPRIT–estimation of signal parameters via rotational invariance techniques. In F. A. Grünbaum, J. W. Helton, and P. Khargonekar, editors, *Signal Processing Part II: Control Theory and Applications*, pages 369–411, New York, 1990. Springer.

[60] E. Schmidt. Zur Theorie der linearen und nichtlinearen Integralgleichungen. I Teil. Entwicklung willkürlichen Funktionen nach System vorgeschriebener. *Mathematische Annalen*, 63:433–476, 1907.

[61] G. N. Stenbakken, T. M. Souders, and G. W. Stewart. Ambiguity groups and testability. *IEEE Transactions on Instrumentation*, 38:941–947, 1989.

[62] G. W. Stewart. The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM Journal on Numerical Analysis*, 17:403–404, 1980.

[63] G. W. Stewart. On the invariance of perturbed null vectors under column scaling. *Numerische Mathematik*, 44:61–65, 1984.

[64] G. W. Stewart. Rank degeneracy. *SIAM Journal on Scientific and Statistical Computing*, 5:403–413, 1984.

[65] G. W. Stewart. Collinearity and least squares regression. *Statistical Science*, 2:68–100, 1987.

[66] G. W. Stewart. Updating a rank-revealing ULV decomposition. Technical Report CS-TR 2627, Department of Computer Science, University of Maryland, 1991.

[67] G. W. Stewart. On the early history of the singular value decomposition. Technical Report CS-TR-2848, Department of Computer Science, University of Maryland, College Park, 1992.

[68] G. W. Stewart. An updating algorithm for subspace tracking. *IEEE Transactions on Signal Processing*, 40:1535–1541, 1992.

[69] G. W. Stewart. Updating URV decompositions in parallel. Technical Report CS-TR-2880, Department of Computer Science, University of Maryland, College Park, 1992.

[70] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory.* Academic Press, Boston, 1990.

[71] A. N. Tihonov. Regularization of incorrectly posed problems. *Soviet Mathematics*, 4:1624–1627, 1963.

[72] S. Van Huffel. *Analysis of the Total Least Squares Problem and Its Use in Parameter Estimation.* PhD thesis, Katholeike Universiteit Leuven, 1987.

[73] S. Van Huffel and J. Vandervalle. Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in A are subject to error. *SIAM Journal on Matrix Analysis and Applications*, 10:294–315, 1989.

[74] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis.* SIAM, Philadelphia, 1991.

[75] J. M. Varah. A practical examination of some numerical methods for linear discrete ill-posed problems. *SIAM Review*, 21:100–111, 1979.

[76] J. H. Wilkinson and C. Reinsch. *Handbook for Automatic Computation. Vol. II Linear Algebra.* Springer, New York, 1971.

[77] G. Xu and T. Kailath. Fast signal-subspace decomposition — Part I: Ideal covariance matrices. Manuscript submitted to ASSP. Information Systems Laboratory, Stanford University, 1990.

[78] G. Xu and T. Kailath. Fast signal-subspace decomposition — Part II: Sample covariance matrices. Manuscript submitted to ASSP. Information Systems Laboratory, Stanford University, 1990.