University of Maryland College Park

Institute for Advanced Computer Studies TR–97–75
Department of Computer Science TR–3840

On an Inexpensive Triangular Approximation to the Singular Value Decomposition*

G. W. Stewart†

ABSTRACT

In this paper we introduce a new decomposition called the pivoted QLP decomposition. It is computed by applying pivoted orthogonal triangularization to the columns of the matrix X in question to get an upper triangular factor R and then applying the same procedure to the rows of R to get a lower triangular matrix L. The diagonal elements of R are called the R-values of X; those of L are called the L-values. Numerical examples show that the L-values track the singular values of X with considerable fidelity — far better than the R-values. At a gap in the L-values the decomposition provides orthonormal bases of analogues of row, column, and null spaces provided of X. The decomposition requires no more than twice the work required for a pivoted QR decomposition. The computation of R and L can be interleaved, so that the computation can be terminated at any suitable point, which makes the decomposition especially suitable for low-rank determination problems. The interleaved algorithm also suggests a new, efficient 2-norm estimator.

*This report is an extensive revision and expansion of TR–97–31, Department of Computer Science, University of Maryland. Both are available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports` or on the web at `http://www.cs.umd.edu/ stewart/`.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported by the National Science Foundation under grant CCR-95-03126.

On an Inexpensive Triangular Approximation to the Singular Value Decomposition*

G. W. Stewart†

Abstract

In this paper we introduce a new decomposition called the pivoted QLP decomposition. It is computed by applying pivoted orthogonal triangularization to the columns of the matrix X in question to get an upper triangular factor R and then applying the same procedure to the rows of R to get a lower triangular matrix L. The diagonal elements of R are called the R-values of X; those of L are called the L-values. Numerical examples show that the L-values track the singular values of X with considerable fidelity — far better than the R-values. At a gap in the L-values the decomposition provides orthonormal bases of analogues of row, column, and null spaces provided of X. The decomposition requires no more than twice the work required for a pivoted QR decomposition. The computation of R and L can be interleaved, so that the computation can be terminated at any suitable point, which makes the decomposition especially suitable for low-rank determination problems. The interleaved algorithm also suggests a new, efficient 2-norm estimator.

1. Introduction

This paper concerns the problem of locating gaps in the singular values of a matrix. Specifically, suppose that X is an $n \times p$ matrix with $n \geq p$. Then there are orthogonal matrices U and V such that

$$U^{\mathrm{T}} X V = \begin{pmatrix} \Sigma \\ 0 \end{pmatrix}, \tag{1.1}$$

where

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p), \qquad \sigma_1 \geq \cdots \geq \sigma_p.$$

The decomposition (1.1) is called the singular value decomposition of X, and the scalars σ_i are called singular values. We say that X has a (relative) gap at m if σ_{m+1}/σ_m is small.

A gap in the singular values may reflect an underlying rank degeneracy in a matrix of which X is a perturbation. Or it may simply be a convenient point at which to reduce the dimensionality of the problem represented by X. In either case one is usually

*This report is an extensive revision and expansion of TR–97–31, Department of Computer Science, University of Maryland.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported by the National Science Foundation under grant CCR-95-03126.

interested in certain subspaces associated with the decomposition. Suppose we partition the decomposition in the form

$$\begin{pmatrix} U_1^{\mathrm{T}} \\ U_2^{\mathrm{T}} \end{pmatrix} X(V_1 \;\; V_2) = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{pmatrix},$$

where Σ_1 is of order m. If $\Sigma_2 = 0$, then the column spaces of U_1 and V_1 are the column spaces of X and X^{T}, while the column spaces of U_1 and V_1 are the null spaces of X and X^{T}. When Σ_2 is merely small, the columns of U and V provide useful analogues of these spaces. To fix our nomenclature, we will call

$\mathcal{R}(U_1)$	the right superior singular subspace,
$\mathcal{R}(U_2)$	the right inferior singular subspace,
$\mathcal{R}(V_1)$	the left superior singular subspace,
$\mathcal{R}(V_2)$	the left inferior singular subspace,

and collectively call the spaces *fundamental spaces.*[1]

Since gaps and singular subspaces are defined in terms of the singular value decomposition, the natural way to compute them is to compute the singular value decomposition. Unfortunately, the computation of this decomposition is expensive. For this reason, researchers have proposed many alternatives. Of these the pivoted QR decomposition is widely recommended because of its simplicity. As we shall see, however, the pivoted QR decomposition has certain drawbacks: it gives only fuzzy approximations to the singular values, and it fails to provide orthonormal bases for some of the fundamental subspaces. The purpose of this paper is to investigate the consequences of the empirical observation that if we reduce the R-factor of the pivoted QR decomposition to lower triangular form the diagonals track the singular values with considerable fidelity. We call this decomposition the pivoted QLP decomposition.

The paper is organized as follows. In Section 2 we introduce the pivoted QR decomposition and discuss its properties. In Section 3 we introduce the pivoted PLQ decomposition and present some numerical experiments that show its abilities to track singular values. Section 4 is devoted to the approximations to the fundamental subspaces provided by the new decomposition. In Section 5 we discuss implementation issues. It turns out that the computation of the R factor and the L factor can be interleaved, making the decomposition efficient for low rank problems. The interleaving can also be used to derive an efficient 2-norm estimator, which is the subject of Section 6.

[1] Gill Strang [11] introduced the modifier "fundamental" to describe the column, row, and null spaces of a matrix. Per Christian Hansen extended the usage to their analogues. The use of "superior" and "inferior" in this connection is new.

The paper concludes with a summary and a discussion of open problems associated with the decomposition.

Throughout this paper, $\|\cdot\|$ will denote the Euclidean vector norm and the subordinate spectral matrix norm. The smallest singular value of X will be denoted by $\inf(X)$.

2. The pivoted QR decomposition

As above let X be an $n \times p$ matrix with $n \geq p$. Then for any permutation matrix Π_{R} there is an orthogonal matrix Q such that

$$Q^{\mathrm{T}} X \Pi_{\mathrm{R}} = \begin{pmatrix} R \\ 0 \end{pmatrix}, \tag{2.1}$$

where R is upper triangular. The matrix Π can be chosen so that the elements of R satisfy

$$r_{kk}^2 \geq \sum_{i=k}^{p} r_{ij}^2, \qquad j = k+1, \ldots, p.$$

In other words if R_{kk} denotes the trailing submatrix of R of order $p-k+1$, then the norm of the first column of R_{kk} dominates the norms of the other columns. We will call this decomposition the *pivoted* QR *decomposition*.

The pivoted QR decomposition was introduced by Golub [6], who computed it by a variant of Householder's orthogonal triangularization. Specifically, at the kth stage of the reduction, we will have computed Householder transformations $H_1, \ldots, H_{k-1}$ and permutations $\Pi_1, \ldots, \Pi_{k-1}$ such that

$$H_{k-1} \cdots H_1 X \Pi_1 \cdots \Pi_{k-1} = \begin{pmatrix} R_{11} & R_{12} \\ 0 & X_k \end{pmatrix}, \tag{2.2}$$

where R_{11} is upper triangular. Before the kth Householder transformation H_k is computed, the column of X_k of largest norm (along with the corresponding column of R_{12}) is swapped with the kth column. Since the norm of this column is $|r_{kk}|$, the pivoting strategy can be regarded as a greedy algorithm to make the leading principal submatrices of R as well conditioned as possible by making their trailing diagonals as large as possible.

We will call the diagonals of R the R-*values* of X. The folklore has it that the R-values track the singular values well enough to expose gaps in the latter. For example, a matrix X of order 100 was generated in the form

$$X = U \Sigma V^{\mathrm{T}} + 0.1 \sigma_{50} E,$$

where

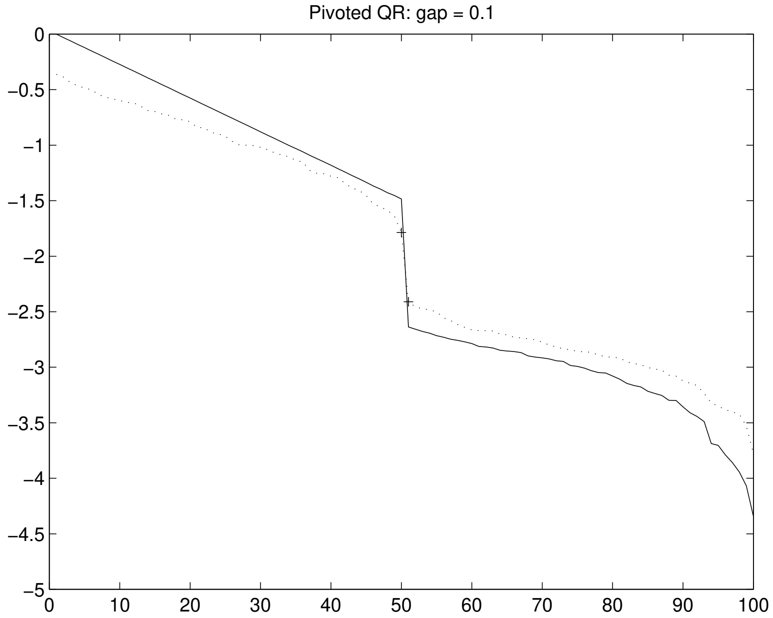

Figure 2.1: Gap reveliation in pivoted QR

1. Σ is formed by creating a diagonal matrix with diagonals decreasing geometrically from one to 10^{-3} and setting the last fifty diagonal elements to zero,
2. U and V are random orthogonal matrices,
3. E is a matrix of standard normal deviates.

Thus X represents a matrix of rank 50 perturbed by an error whose elements are one-tenth the size of the last nonzero singular value. Figure 2.1 plots the common logarithms of the singular values of X (continuous line) and R-values of X (dotted line) against their indices. The +'s indicate the values of $r_{50,50}$ and $r_{51,51}$. It is seen that there is a well-marked gap in the R-values, though not as marked as the gap in the singular values.

If we we detect a gap at m and partition

$$\hat{X} \equiv X\Pi_{\mathrm{R}} = (Q_1\ \ Q_2\ \ Q_\perp)\begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{pmatrix}, \tag{2.3}$$

where R_{11} is $m\times n$, then Q_1 and $(Q_2\ Q_\perp)$ are approximate the the right fundamental subspaces of X. Perhaps more important, if we partition

$$\hat{X} = (\hat{X}_1\ \hat{X}_2), \tag{2.4}$$

it follows that $\hat{X}_1 = R_{11}Q_1$ and hence that $\hat{X}_1$ is a basis for the left superior subspace of X. Thus the pivoted QR decomposition extracts a natural basis from the among the columns of X. In applications where these columns represent variables in a model, this basis is a conceptual as well as a computational economization.

The decomposition also provides bases

$$\Pi_{\mathrm{R}} \begin{pmatrix} R_{11}^{\mathrm{T}} \\ R_{12}^{\mathrm{T}} \end{pmatrix} \quad \text{and} \quad \Pi_{\mathrm{R}} \begin{pmatrix} I \\ -R_{11}^{-1}R_{12} \end{pmatrix}$$

that approximate the right fundamental subspaces. Unfortunately, these bases are not orthonormal, and the basis for the right inferior subspace requires additional computation.

Although the R-values tend to reveal gaps in the singular values, they can fail spectacularly, as the following example due to Kahan shows. Let K_n be the upper triangular matrix illustrated below for $n = 6$:

$$K_6 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & s & 0 & 0 & 0 & 0 \\ 0 & 0 & s^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & s^3 & 0 & 0 \\ 0 & 0 & 0 & 0 & s^4 & 0 \\ 0 & 0 & 0 & 0 & 0 & s^5 \end{pmatrix} \begin{pmatrix} 1 & -c & -c & -c & -c & -c \\ 0 & 1 & -c & -c & -c & -c \\ 0 & 0 & 1 & -c & -c & -c \\ 0 & 0 & 0 & 1 & -c & -c \\ 0 & 0 & 0 & 0 & 1 & -c \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Here $c^2 + s^2 = 1$. All the columns of the matrix have the same 2-norm — namely one — so that if ties in pivoting process are broken by choosing the first candidate, the first step of pivoted orthogonal triangularization leaves the matrix unchanged. Similarly for the the remaining steps. Thus pivoted orthogonal triangularization leaves K_n unchanged, and the smallest R-value is s^{n-1}.

However, the matrix can have singular values far smaller than s^{n-1}. The following table

c	σ_{99}	σ_{100}	$r_{99,99}$	$r_{100,100}$
0.0	1.0e+00	1.0e+00	1.0e+00	1.0e+00
0.1	6.4e−01	9.5e−05	6.1e−01	6.1e−01
0.2	1.5e−01	3.7e−09	1.4e−01	1.3e−01
0.3	1.1e−02	9.3e−14	9.8e−03	9.4e−03
0.4	2.3e−04	1.1e−18	1.9e−04	1.8e−04

(2.5)

presents the 99th and 100th singular and R-values of K_{100} for various values of c. When $c = 0$, $K_n = I$, and the R-values and singular values coincide. As c departs from zero, however, there is an increasingly great gap between the next-to-last and last singular values, while the ratio of the corresponding R-values remains near one.

This example has inspired researchers to look for other pivoting strategies under the rubric of rank revealing QR decompositions [2, 3, 4, 5, 7, 8]. There are, however, certain limitations to *any* pivoted QR decomposition. For example, the first R-value is the norm of the first column of $A\Pi_{\mathrm{R}}$. We hope this number will approximate σ_1, which, however, is the spectral norm of the entire matrix A. Thus r_{11} will in general underestimate σ_1 (this fact also follows from the minimax characterization of singular values). Similarly, $|r_{pp}|$ will generally overestimate σ_p.

For example, consider the rank-one matrix $X = \mathbf{e}\mathbf{e}^{\mathrm{T}}$, where $\mathbf{e}$ is the vector whose components are all one. The norm of this matrix is $\sqrt{np}$. On the other hand, the columns of X all have norm $\sqrt{n}$. Thus if $n = p = 100$, the first R-value underestimates the corresponding singular value by a factor of 10 — regardless of the pivoting strategy. Figure 2.1 also illustrates this tendency of the pivoted QR decomposition to underestimate the largest singular value. In fact, although the R-values reveal the gap, they do not give a very precise representation of the distribution of the singular values. We now turn to a postprocessing step that seems to clean up these values.

3. The pivoted QLP decomposition

To motivate our new decomposition consider the partitioned R-factor

$$R = \begin{pmatrix} r_{11} & r_{12}^{\mathrm{T}} \\ 0 & R_{22} \end{pmatrix}$$

of the pivoted QR decomposition. We have observed that r_{11} is an underestimate of $\|X\|_2$. A better estimate is the norm $\ell_{11} = \sqrt{r_{11}^2 + r_{12}^{\mathrm{T}} r_{12}}$ of the first row of R. We can calculate that norm by postmultiplying R by a Householder transformation H_1 that reduces the first row of R to a multiple of $\mathbf{e}_1$:

$$RH_1 = \begin{pmatrix} \ell_{11} & 0 \\ \ell_{12} & \hat{R}_{22} \end{pmatrix}.$$

We can obtain an even better value if we interchange the largest row of R with the first:

$$\Pi_1 R H_1 = \begin{pmatrix} \ell_{11} & 0 \\ \ell_{12} & \hat{R}_{22} \end{pmatrix}. \tag{3.1}$$

Now if we transpose (3.1), we see that it is the first step of pivoted Householder triangularization applied to R^{T} [cf. (2.2)]. If we continue this reduction and transpose

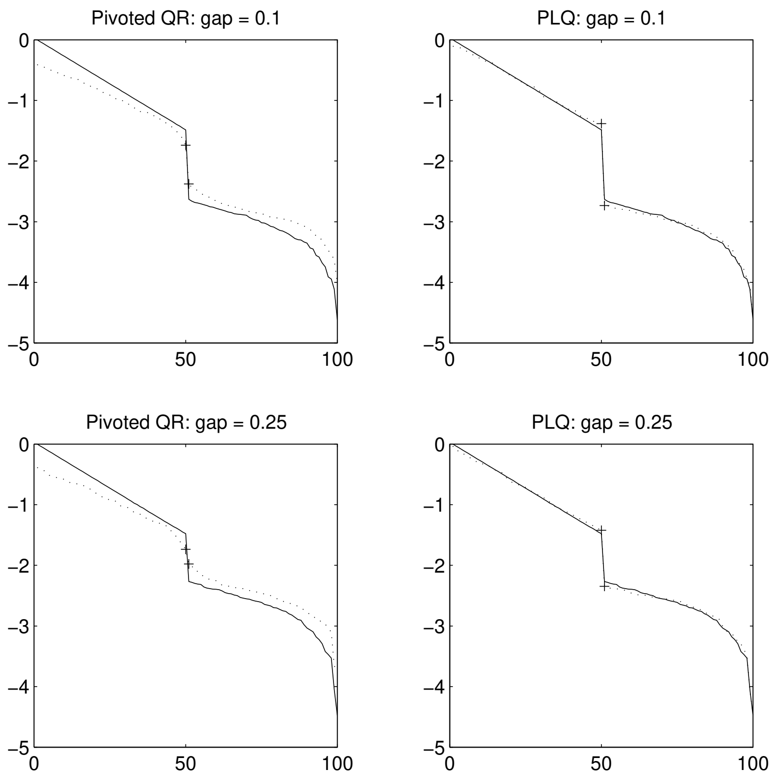

Figure 3.1: Pivoted QR and QLP decompositions compared

the result, we obtain a triangular decomposition of the form

$$\Pi_{\mathrm{L}}^{\mathrm{T}} Q^{\mathrm{T}} X \Pi_{\mathrm{R}} P = \begin{pmatrix} L \\ 0 \end{pmatrix}.$$

We will call this the *pivoted QLP decomposition* of X and will call the diagonal elements of L the *L-values* of X.[2]

The way we motivated the pivoted QLP decomposition suggests that it might provide better approximations to the singular values of the original matrix X than does the pivoted QR decomposition. The top two graphs in Figure 3.1 compare performance of the two decompositions on the matrix of Figure 2.1. The solid lines, as above, indicate

[2]The decomposition has been introduced independently by Hosoda [9], who uses it to regularize ill-posed problems.

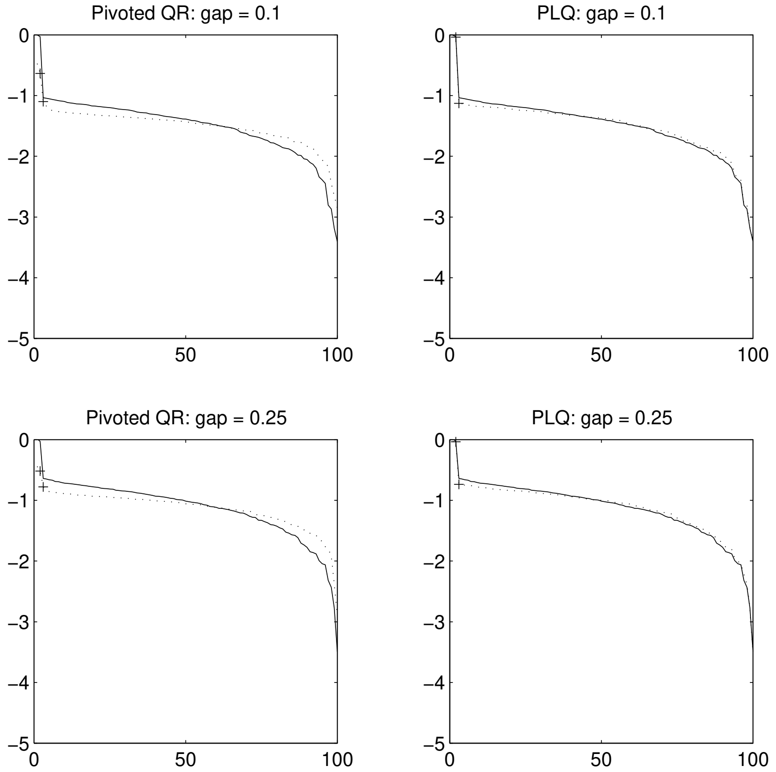

Figure 3.2: Rank determination: Low rank

singular values and the dotted lines represent R-values on the left and L-values on the right. It is seen that that in comparison with the R-values, the L-values track the singular values with remarkable fidelity.

The lower pair of graphs shows the behavior of the decomposition when the gap is reduced from a ratio of 0.1 to 0.25. Here the L-values perform essentially as well as the singular values. The gap in the R-values is reduced to the point where an automatic gap-detecting algorithm might fail to see it.

The superiority becomes even more marked in the low-rank case illustrated in Figure 3.2. The R-values only suggest the gap of 0.25 and reveal the gap of 0.1 weakly. The L-values are as good as one could hope for.

Figure 3.3 presents a more taxing example — called the devil's stairs — in which the singular values have multiple gaps. When the gaps are small, as in the top two

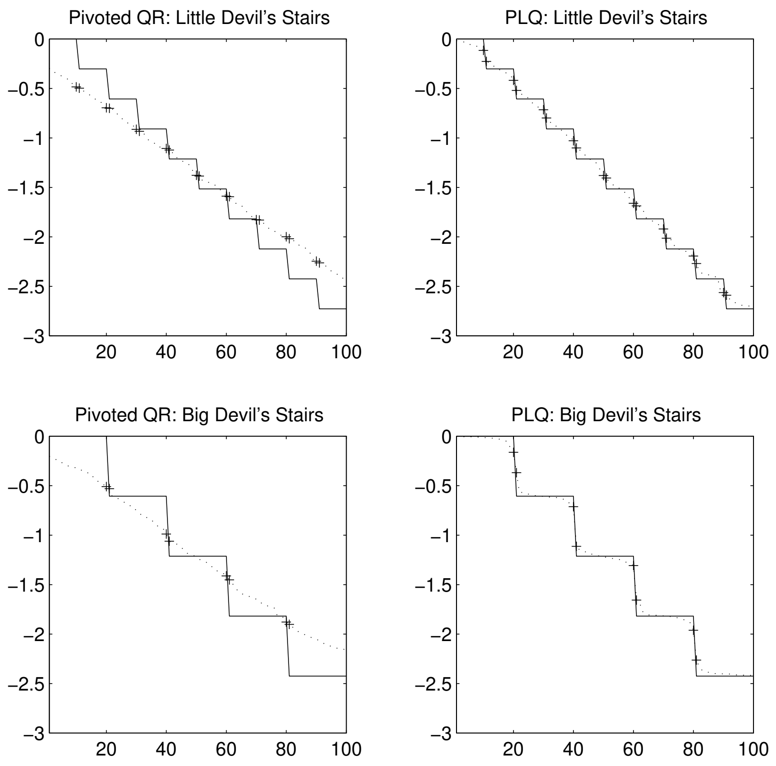

Figure 3.3: The devil's stairs

graphs, neither decomposition does well at exhibiting their presence, although the L-values track the general trend of the singular values far better than the R-values. In the pair of graphs at the bottom, the gaps are fewer and bigger. Here the L-values clearly reveal the gaps, while the R-values do not.

The pivoted QLP decomposition takes Kahan's example in stride. The following table — the QLP analogue of (2.5) — shows that the last two L-values are good approximations of the corresponding singular values.

c	σ_{99}	σ_{100}	$\ell_{99,99}$	$\ell_{100,100}$
0.0	1.0e+00	1.0e+00	−1.0e+00	1.0e+00
0.1	6.4e−01	9.5e−05	−4.8e−01	2.2e−04
0.2	1.5e−01	3.7e−09	−1.1e−01	6.4e−09
0.3	1.1e−02	9.3e−14	−9.0e−03	1.4e−13
0.4	2.3e−04	1.1e−18	−1.9e−04	1.5e−18

The above examples suggest that not only does the pivoted QLP decomposition reveal gaps in the singular values better than the pivoted QR decomposition but that it is also superior in tracking the singular values. Pivoting plays a key role in this. Pivoting in the QR stage of the algorithm is absolute necessary. For the above examples, the pivoting in the reduction to lower triangular form is largely cosmetic — it enforces monotonicity of the L-values. Without pivoting the L values tend to behave like their pivoted counterparts. Nonetheless, pivoting at this stage may be necessary. For example, consider the following matrix of order n:

$$\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{n}}\mathbf{e}\mathbf{e}^{\mathrm{T}} \end{pmatrix}.$$

Without pivoting the first L value is one, which underestimates the norm $(n-1)/\sqrt{n}$. On the other hand, the pivoted L value reproduces the norm exactly.

4. Fundamental subspaces

If we incorporate the pivots in the QLP decomposition into the orthogonal transformations by defining

$$\hat{Q} = Q\Pi_{\mathrm{L}} \quad \text{and} \quad \hat{P} = \Pi_{\mathrm{R}}P,$$

then the decomposition can be written in the partition form

$$X = (\hat{Q}_1 \ \hat{Q}_2 \ Q_\perp) \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{P}_1^{\mathrm{T}} \\ \hat{P}_2^{\mathrm{T}} \end{pmatrix}, \tag{4.1}$$

where L_{11} is of order m. Since the L-values tend to track the singular values, if there is a gap in the latter at m, the partition of $\hat{P}$ and $\hat{Q}$ provide orthonormal bases approximating the four fundamental subspaces of X at m. Specifically,

1. $\mathcal{R}(\hat{Q}_1)$ approximates the left superior subspace of X,
2. $\mathcal{R}[(\hat{Q}_2 \ Q_\perp)]$ approximates the left inferior subspace of X,

3. $\mathcal{R}(\hat{P}_1)$ approximates the right superior subspace of X,

4. $\mathcal{R}(\hat{P}_2)$ approximates the right inferior subspace of X.

Thus the pivoted QLP decomposition, like the pivoted QR decomposition, furnishes orthonormal approximations to the left fundamental subspaces, but unlike the latter, it also furnishes orthonormal approximations to the right fundamental subspaces.

A nice feature of the the decomposition is that the accuracy of the approximate subspaces can be estimated from the the decomposition itself. We will measure the concordance of two subspaces $\mathcal{X}$ and $\mathcal{Y}$ by the sine of their largest canonical angle between them. This number is the largest singular value of $X^{\mathrm{T}}Y_{\perp}$, where X is an orthonormal basis for $\mathcal{X}$ and $Y_{\perp}$ is an orthonormal basis for the orthogonal complement of $\mathcal{Y}$. The following theorem follows from results in [10].

Theorem 4.1. *Let s_{u} be the sine of the largest canonical angle between the left superior subspace of X and $\mathcal{R}(\hat{Q}_1)$, and let s_{v} be the sine of the largest canonical angle between the right superior subspace of X and $\mathcal{R}(\hat{P}_1)$. If*

$$\rho = \frac{\|L_{22}\|_2}{\inf(L_{11})} < 1,$$

then

$$s_{\mathrm{u}} \le \frac{1}{1-\rho^2}\,\frac{\|L_{21}\|}{\inf(L_{11})} \quad \text{and} \quad s_{\mathrm{v}} \le \frac{\rho}{1-\rho^2}\,\frac{\|L_{21}\|}{\inf(L_{11})} \tag{4.2}$$

The bounds for the left and right inferior subspaces are the same.

Although this theorem is cast in terms of quantities that are difficult to compute, $\inf(L_{11})$ can be estimated by the L-value ℓ_{mm}, while $\|L_{22}\|$ can be estimated by $\ell_{m+1,m+1}$. The quantity $\|L_{21}\|$ can be bounded by the Frobenius norm or can be estimated by the two norm estimator described in Section 6.

In Section 2 we stressed the desirability of having the left superior subspace at a gap associated with a specific set of columns of X. For the pivoted QR decomposition we showed that the columns of $\hat{X}_1$ in (2.4) spanned the same space as those of Q_1 in 2.3, and hence $\hat{X}_1$ provides such a basis. In the pivoted QRP decomposition we must replace Q by

$$\hat{Q} \equiv Q\Pi_{\mathrm{L}} = (\hat{Q}_1 \ \hat{Q}_2 \ Q_{\perp}).$$

In general, the pivoting in the reduction to lower triangular form mixes up the columns of Q so that $\hat{Q}_1$ cannot be associated with a set of columns of X. However, if the partition (4.1) corresponds to a substantial gap in the R-values, it is unlikely that the pivoting process will interchange columns across Q_1 and Q_2. In this case the column spaces of Q_1 and $\hat{Q}_1$ are the same and are spanned by the columns of $\hat{X}_1$. Thus in the applications we are most interested in, the left superior subspace will be associated with a specific set of columns of X.

5. Implementation issues

An advantage of the QLP decomposition is that it can be computed with off the shelf software. The standard matrix packages have programs for computing a pivoted QR decomposition. To compute the QLP decomposition one merely applies the program twice, once to X and once to R^{T}.

Householder triangularization requires about $np^2 - \frac{1}{3}p^3$ floating point additions and multiplications. The additional work in the reduction to lower triangular form is $\frac{2}{2}p^3$ additions and multiplications.. Thus when $n = p$ the pivoted QLP decomposition requires twice as much work as the pivoted QR decomposition. As n increases, the additional work becomes negligible.

It should be pointed out that when n is much greater than p the work required for calculation of the singular value decomposition of R (and hence that of X) is also negligible (e.g., see [1]). However, an important aspect of the QLP decomposition — one not shared with the singular value decomposition — is that by interleaving the computation of R and L we can terminate the algorithm as soon as it has proceeded far enough for the purposes at hand. Specifically, at the kth step of the reduction to upper triangular form, we have

$$H_{k-1} \cdots H_1 X \Pi_1 \cdots \Pi_{k-1} = \begin{pmatrix} R_{11} & R_{1k} \\ 0 & X_{kk} \end{pmatrix}.$$

Since the first $k-1$ rows of R are present in this decomposition, we can reduce them to get first $k-1$ rows of L. We can then compute more rows of R, and reduce them to get the corresponding of rows of L. And so on. This process restricts the pivoting in forming L to the set of rows currently being reduced. However, as we have noted, except for contrived examples, pivoting in the formation of L does not much improve the L-values.

This interleaved pivoting is especially valuable in low-rank problems, since one can stop computing the decomposition after a gap has been found, with a potentially great savings in work. In fact, since the gap revealing abilities of the R-values are themselves not negligible, one can treat the computation of R as a probe for potential gaps, which can then be confirmed by computing the corresponding rows of L. The interleaved pivoting can also be used to improve the efficiency of Hosoda's regularization algorithm [9].

6. A 2-norm estimator

We have see that the first L-value in the pivoted QLP decomposition of a matrix X is generally a good estimate of $\|X\|_2$. Since we can interleave the computation of R and L, we can estimate $\|X\|_2$ by computing a few rows of the pivoted R-factor. Usually (but not always) the norm of the largest row will be the $(1,1)$-element of the pivoted L-factor — the first L value. In this section we will describe how to implement this estimation scheme.

The basic idea is that the pivoted R-factor of X is the pivoted Cholesky factor of the cross-product matrix $A = X^{\mathrm{T}}X$. We will now show how to compute the factor row by row. We begin with the unpivoted version.

Suppose we have computed the first $k-1$ rows of R and partition them in the form

$$(R_{11} \;\; R_{12}),$$

where R_{11} is triangular of order $k-1$. Then

$$A - \begin{pmatrix} R_{11}^{\mathrm{T}} \\ R_{12}^{\mathrm{T}} \end{pmatrix} (R_{11} \;\; R_{12}) = \begin{pmatrix} 0 & 0 \\ 0 & S \end{pmatrix},$$

where S is the Schur complement of $A_{11} = R_{11}^{\mathrm{T}} R_{11}$ in A. Since the kth row of R is the first row of S divided by the square root of the $(1,1)$-element of S, we can compute the elements of the kth row in a Crout-like algorithm as follows.

1. **for** $j = k$ **to** p
2. $\quad r_{kj} = x_k^{\mathrm{T}} x_j$
3. $\quad r_{kj} = r_{kj} - r_{1k}r_{1j} - \cdots - r_{k-1,k}r_{k-1,j}$
4. $\quad r_{kj} = r_{kj}/\sqrt{r_{kk}}$
5. **end for** k

The first statement in the loop generates an element in the kth row of A. The second statement computes the Schur complement of that element. The third statement scales the element so that it becomes an element of the R-factor.

Turning now to the pivoted algorithm, we must answer two questions.

1. How do we keep track of the norms to determine the pivots?

2. How do we organize the interchanges?

The answer to the first question is that the squares of the norms that determine the pivots are the diagonals of the Schur complement. These quantities can be formed initially and downdated as we add rows. The answer to the second question is that we perform no interchanges. Instead we keep track of indices of the columns we have selected as pivots and skip operations involving them as we add the kth row. For example, with a pivot sequence of $3, 2, 4, 1$ in a 4×4 matrix we would obtain an "R-factor" of the form

$$\begin{pmatrix} r_{14} & r_{12} & r_{11} & r_{13} \\ r_{24} & r_{22} & 0 & r_{23} \\ r_{34} & 0 & 0 & r_{33} \\ r_{44} & 0 & 0 & 0 \end{pmatrix}.$$

Given an $n{\times}p$ matrix X, this algorithm returns and estimate *norm2est* of $\|X\|_2$.

```
normx2[j] = ||X[:, j]||_2^2  (j = 1, ..., p)
P = ∅
norm2est = 0
for k = 1 to kmax
    Choose pvt ∉ P so that nrmx2[pvt] ≥ nrmx2[j]  (j ∉ P)
    P = P ∪ {pvt}
    nr2 = nrmx2[pvt]
    rkk = sqrt(nr2)
    for j = 1 to p, j ∉ P
        R[k, j] = X[:, pvt]^T*X[:, j]
        for i = 1 to k−1
            R[k, j] = R[k, j] − R[i, pvt]*R[i, j]
        end for i
        R[k, j] = R[k, j]/rkk
        nr2 = nr2 + R[k, j]^2
        normx2[j] = normx2[j] − R[k, j]^2
    end for j
    norm2est = max{norm2est, nr2}
end for k
norm2est = sqrt(norm2est)
```

Figure 6.1: A 2-norm estimator

Figure 6.1 contains an implementation of this scheme. The heart of it is the the loop on j, in which $R[k, j]$ is initialized, transformed into the corresponding element of the Schur complement, and normalized. The norms of the reduced columns of X, which are contained in the array *normx2* are also updated at this point. The square of the norm of the row is accumulated in *nr2* and then compared with *norm2est*. Here are some comments.

1. The variable *kmax* is an upper bound on the number of rows of R to compute. Two or perhaps three is a is a reasonable number. Another possibility is to compute rows of R until a decrease in norm occurs.

2. The bulk of the work is in computing the norms of the columns of X and initializing R. Specifically if X is $n{\times}p$ and *kmax* is not too large, the algorithm requires about $(kmax{+}1)np$ additions and multiplications.

3. The algorithm can produce an underestimate. For example, consider the $n \times n$ matrix
$$X = \begin{pmatrix} I_{kmax} & 0 \\ 0 & \frac{1}{\sqrt{n}}\mathbf{e}\mathbf{e}^{\mathrm{T}} \end{pmatrix}.$$
Because the first *kmax* columns of X dominate the rest, the estimator will return a value of *norm2est* of one. But the norm of the matrix is $(n - kmax)/\sqrt{n}$.

7. Discussion

The pivoted QLP decomposition of an $n \times p$ matrix X is computed by applying pivoted orthogonal triangularization to the columns of X to get an upper triangular factor R and then applying the same procedure to the rows of R to get a lower triangular matrix L. The diagonal elements of R are called the R-values of X; those of L are called the L-values. This decomposition has several attractive properties.

1. The L-values track the singular values of X with considerable fidelity — far better than the R-values. This makes the decomposition particularly appropriate for rank determination.

2. The decomposition provides orthonormal bases for approximations to the four fundamental subspaces from the singular value decomposition. The accuracy of these approximations can be estimated from the decomposition itself. When the decomposition is divided by at a gap, it provides a well-conditioned subset of the columns of X that span the left superior subspace.

3. The decomposition requires no more than twice the work required for a pivoted QR decomposition. The additional work decreases as n increases.

4. The computation of R and L can be interleaved. This makes the decomposition suitable for low-rank problems. It can also be used to regularize systems resulting from ill-posed problems.

5. The decomposition suggests an efficient method for estimating the 2-norm of a matrix.

Why the L-values track the singular values so well is an open question. Without pivoting, the decomposition represents the first two steps in an iterative algorithm for computing the singular value decomposition by reducing a matrix successively to upper then lower triangular form. However, the convergence of this algorithm depends on the ratios of neighboring singular values, which in our examples are too near one to account for the behavior we have observed. The effectiveness of the decomposition is intimately tied to the pivoting, the pivoting in the formation of R being essential and the pivoting in the formation of L being necessary to avoid certain contrived counterexamples. I conjecture that the analysis of this decomposition will not be simple.

References

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

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

[3] T. F. Chan and P. C. Hansen. Computing truncated singular value decomposition least squares solutions by rank revealing QR-factorizations. *SIAM Journal on Scientific and Statistical Computing*, 11:519–530, 1990.

[4] T. F. Chan and P. C. Hansen. Some applications of the rank revealing QR factorization. *SIAM Journal on Scientific and Statistical Computing*, 13:727–741, 1992.

[5] S. Chandrasekaran and I. C. F. Ipsen. On rank-revealing factorizations. *SIAM J. Matrix Anal. Appl.*, 15:592–622, 1994.

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

[7] M. Gu and S. C. Eisenstat. An efficient algorithm for computing a strong rank-revealing QR-factorization. *SIAM Journal on Scientific Computing*, 17:848–869, 1996.

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

[9] Y. Hosoda. Anew method for linear ill-posed problems with double use of the QR-decomposition. Paper presented at the Kalamazoo Symposium on Matrix Analysis and Applications, Kalamazoo, Michigan, October, 1997.

[10] R. Mathias and G. W. Stewart. A block QR algorithm and the singular value decomposition. *Linear Algebra and Its Applications*, 182:91–100, 1993.

[11] G. Strang. *Linear Algebra and Its Applications*. Academic Press, New York, third edition, 1988.