UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Stewart's Pivoted QLP Decomposition for Low-Rank Matrices

David A. Huckaby
Tony F. Chan

September 2002

CAM Report 02-54

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90095-1555

http://www.math.ucla.edu/applied/cam/index.html

Stewart's Pivoted QLP Decomposition for Low-Rank Matrices

David A. Huckaby and Tony F. Chan

September 9, 2002

Abstract

The pivoted QLP decomposition, introduced by G. W. Stewart [19], represents the first two steps in an algorithm which approximates the SVD. If A is an m-by-n matrix, the matrix $A\Pi_0$ is first factored as $A\Pi_0 = QR$, and then the matrix $R^T\Pi_1$ is factored as $R^T\Pi_1 = PL^T$, resulting in $A = Q\Pi_1 LP^T\Pi_0^T$, with Q and P orthogonal, L lower-triangular, and Π_0 and Π_1 permutation matrices. The Q and P matrices provided approximations of the left and right singular subspaces, and the diagonal elements of L are excellent approximations of the singular values of A. Stewart observed that pivoting is not necessary in the second step, allowing one to efficiently truncate the decomposition, computing only the first few columns of R andn L and choosing the stopping point dynamically. In this paper, we demonstrate that this truncating actually works by extending our theory for the complete pivoted QLP decomposition [11]. In particular, say there is a gap between σ_k and σ_{k+1}, and partition the matrix L into diagonal blocks L_{11} and L_{22} and off-diagonal block L_{21}, where L_{11} is k-by-k. If we compute only the block L_{11}, the convergence of $(\sigma_j(L_{11})^{-1} - \sigma_j^{-1})/\sigma_j^{-1}$ for $j = 1, \ldots, k$ are all quadratic in the gap ratio σ_{k+1}/σ_k. Hence, if the gap ratio is small as it usually is when A has numerical rank k, then all of the singular values are likely to be well approximated. This truncated pivoted QLP decomposition can be computed in $\mathcal{O}(mnk)$ time, making it ideal for accurate SVD approximations for low-rank problems.

1 Introduction

Many applications involve a large m-by-n matrix A with numerical low rank; that is, with a few significant singular values and the rest close to zero. The Singular Value Decomposition (SVD) of A is given by $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$ is the matrix of left singular vectors, $V \in \mathbb{R}^{n\times n}$ is the matrix of right singular vectors, and $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix containing the singular values. We would like to have a rank k approximation of A which ignores the contributions of the small singular values. It is well known that the best such

approximation to A is obtained by simply truncating the SVD: $A \approx A_k = U_k \Sigma_k V_k^T$, where U_k is m-by-k, Σ_k is k-by-k, and V_k is k-by-n.

Our paper focuses on an approximation to the truncated SVD based on a relatively new matrix decomposition. The pivoted QLP decomposition, introduced by G. W. Stewart [19], can be viewed as an approximate SVD. If A is an m-by-n matrix, the matrix $A\Pi_0$ is first factored as $A\Pi_0 = QR$, and then the matrix $R^T\Pi_1$ is factored as $R^T\Pi_1 = PL^T$, resulting in $A = Q\Pi_1 L P^T \Pi_0^T$, with Q and P orthogonal, L lower-triangular, and Π_0 and Π_1 permutation matrices. Stewart observed that the elements on the diagonal of L tend to be good approximations of the singular values of A, and we [11] provided some theory to explain this. We showed that if there is a substantial gap somewhere in the singular values, then the pivoted QLP decomposition is likely to give good approximations to all of them. We showed this assuming that pivoting is not even performed on the second step.

Stewart pointed out that because pivoting is not crucial on the second step, we can compute a version of the pivoted QLP decomposition which approximates a truncated SVD. In computing this truncated pivoted QLP decomposition, we can adaptively determine the rank r of the matrix, and thus where we need to truncate, and the entire decomposition can be computed in $\mathcal{O}(mnr)$ time. All of this is possible because in computing the first, say, k_1 columns of L, all that is needed are the first k_1 rows of R. But these rows are available after the first k_1 columns of R have been computed. So the entire computation can be stopped after the k_1 rows of R and columns of L have been computed. This makes the computation very cheap. That it can be restarted, performing the same sequence on the next, say, k_2 rows of R and columns of L, allows us to adaptively find a substantial gap in the diagonal elements of L, and hence the rank of the matrix and the stopping point for the computation.

In this paper, we show that the truncated pivoted QLP decomposition for a rank k problem approximates the first k singular values very well. In particular, say there is a gap between σ_k and σ_{k+1}, and partition the matrix L into diagonal blocks L_{11} and L_{22} and off-diagonal block L_{21}, where L_{11} is k-by-k. If we compute only the block L_{11}, the convergence of $(\sigma_j(L_{11})^{-1} - \sigma_j^{-1})/\sigma_j^{-1}$ for $j = 1, \ldots, k$ are all quadratic in the gap ratio σ_{k+1}/σ_k. We also compute the operation count for the truncated pivoted QLP decomposition, showing it to run in $\mathcal{O}(mnk)$ time, making it essentially a quadratic algorithm when the rank k is small.

2 Truncating and Interleaving

2.1 Explanation of the Method

The QLP decomposition was introduced by G. W. Stewart [19], who observed its potency in rank revelation, singular value approximation, and gap revelation. Let us call the diagonal elements in the R matrix of a QR factorization of A the R-values of A. Noting that the R-values are rough approximations of

the singular values, Stewart suggested taking the pivoted QR factorization and then triangularizing on the right, obtaining the factorization $A = Q\Pi_1 L P^T \Pi_0^T$. If we let $\hat{Q} = Q\Pi_1$ and $\hat{P} = \Pi_0 P$, then we have $A = \hat{Q} L \hat{P}^T$, called the QLP decomposition of A. Note that the second step is equivalent to performing a QR factorization on R^T, obtaining $R^T = P^T L^T$. Also note that L is lower triangular. So the decomposition amounts to taking two pivoted QR factorizations and thus factorizing A into the product of an orthogonal matrix, a lower-triangular matrix, and another orthogonal matrix. The diagonal elements of L are called the L-values of A.

Stewart showed empirically that the L-values track the singular values surprisingly well—far better than the R-values. We showed in [11] that if there is a gap between σ_k and σ_{k+1} and L is partitioned as

$$L = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix}, \tag{1}$$

then the convergence of $(\sigma_j(L_{11})^{-1} - \sigma_j^{-1})/\sigma_j^{-1}$ for $j = 1, \ldots, k$, and of $(\sigma_j(L_{22}) - \sigma_{k+j})/\sigma_{k+j}$, for $j = 1, \ldots, n-k$ are all quadratic in the gap ratio σ_{k+1}/σ_k. These results are true even assuming no pivoting in the second QR factorization.

Stewart noticed empirically that the pivoting in the second step was not crucial, and pointed out that the algorithm could therefore be truncated and/or interleaved. Truncating is simply computing only the first k rows of R and columns of L, where k is perhaps the numerical rank of the matrix A, thus producing an approximation to a truncated SVD of A. Note that assuming pivoting is not necessary on the second step, all that is needed to compute the first k columns of L are the first k columns of R.

Interleaving is a method which allows us to dynamically determine what k should be by alternately probing for gaps in the R-values and using the L-values to confirm the gaps. For example, we can compute rows of R until we find a substantial gap in the R-values, say after computing k_1 rows of R. We can now proceed to compute the first k_1 columns of L. If the suspected gap is verified by the L-values, we are finished. If not, we go back and compute more rows of R until a gap again appears, say after k_2 more rows, then compute the corresponding k_2 columns of L and see whether the gap is verfied. And so forth.

When interleaving, note that within each set of k_i rows of R (columns of R^T) that are used to compute k_i columns of L, pivoting is certainly possible. Note also that when computing the k_i columns of L in the i-th pass, we need the orthogonal transformations (presumably Householder transformations) that upper-triangularized all the previous columns of L. So when interleaving in problems where the number of rows of R and then columns of L being computed in one pass (before going back to compute more rows of R) is rather large, or in problems when many passes are required, it might be advantageous to store the orthogonal transformations in block form (see, for example, [1] and [14]) so that they can be more efficiently applied in the next pass.

Finally, note that there are two sets of orthogonal transformations, those

that upper-triangularize the k computed columns and rows of R (i.e., the first k columns of Q, Q_k) and those that upper-triangularize the k computed columns of L (i.e., the first k columns of P, P_k). The columns of Q_k provide a basis for the left superior singular subspace, while the columns of P_k provide a basis for the right superior singular subspace. So we have approximations to all three matrices in the truncated SVD: $Q_k \approx U_k$, $L \approx \Sigma_k$, and $P_k \approx V_k^T$, where L is k-by-k here. Stewart [19] provides bounds on how well U_k and V_k are approximated by Q_k and P_k, respectively. We next turn to how well L approximates Σ_k.

2.2 Accuracy

We now assess the ability of the truncated pivoted QLP decomposition to capture the singular values. If the full pivoted QLP decomposition is $A = QLP^T$, let L be partitioned as in equation (1), and let R be similarly partitioned as

$$R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}.$$

(Here R_{11} and L_{11} are both k-by-k.) We make the assumption that the initial pivoting reveals the rank in the sense that

$$\|R_{22}\| \leq \sqrt{(k+1)(n-k)}\,\sigma_{k+1}, \tag{2}$$

$$\inf(R_{11}) \geq \frac{\sigma_k}{\sqrt{k(n-k+1)}}. \tag{3}$$

The bounds (2) and (3) are for an RRQR algorithm by Chandrasekaran and Ipsen [3]. (There is a better bound than (3) in Lemma 2.1 of [10].) The column-pivoted QR factorization does not guarantee these bounds but usually delivers them or comes close. We assume these bounds because we are trying to show why the pivoted QLP decomposition usualy works so well. In the same spirit, we also make the assumption that $\rho = \frac{\|R_{22}\|}{\inf(R_{11})} < 1$, which is almost always true if the gap between σ_k and σ_{k+1} is substantial.

We have the following theorem.

Theorem 2.1 *Let A be an m-by-n matrix, $m \geq n$, and let $\sigma_k(A) > \sigma_{k+1}(A)$, $k < n$. Let $(R_{11}R_{12})$ be the first k rows of the R-factor in the pivoted QR factorization of A. Let L_{11}^T be the R-factor in the unpivoted QR factorization of $(R_{11}^T R_{12}^T)^T$. Assume that the bounds (2) and (3) hold and that $\rho = \|R_{22}\|/\inf(R_{11}) < 1$.*

Then for $j = 1, \ldots, k$,

$$\frac{\sigma_j(L_{11})^{-1} - \sigma_j^{-1}}{\sigma_j^{-1}} \leq \left(\frac{\sigma_{k+1}}{\sigma_k}\right)^2 \mathcal{O}\left(\frac{n^{\frac{5}{2}}\|R_{12}\|^2}{\left[1 - \left(\frac{\|L_{22}\|}{\inf(L_{11})}\right)^2\right][\inf(L_{11})]^2}\right). \quad (4)$$

Proof:

Since $\rho = \|R_{22}\|/\inf(R_{11}) < 1$, by Theorem 2.1 of [13] we have $\|L_{22}\| \leq \|R_{22}\|$ and $\inf(L_{11}) \geq \inf(R_{11})$, so that $\rho' = \|L_{22}\|/\inf(L_{11}) < 1$. Hence for $j = 1, \ldots, k$, Theorem 3.1 of [13] gives

$$\frac{\sigma_j(L_{11})}{\sigma_j} \geq 1 - \mathcal{O}\left(\frac{\|L_{12}\|^2}{[1-(\rho')^2][\inf(L_{11})]^2}\right),$$

or

$$\sigma_j(L_{11})^{-1} - \sigma_j^{-1} \leq \sigma_j(L_{11})^{-1}\mathcal{O}\left(\frac{\|L_{12}\|^2}{[1-(\rho')^2][\inf(L_{11})]^2}\right).$$

From this, the result follows:

$$\begin{aligned}
\sigma_j(L_{11})^{-1} - \sigma_j^{-1} &\leq \sigma_j(L_{11})^{-1}\mathcal{O}\left(\frac{\|L_{12}\|^2}{[1-(\rho')^2][\inf(L_{11})]^2}\right) \\
&\leq \sigma_j(L_{11})^{-1}\mathcal{O}\left(\frac{\rho^2\|R_{12}\|^2}{[1-(\rho')^2][\inf(L_{11})]^2}\right) \quad (5) \\
&\leq \sigma_j(R_{11})^{-1}\mathcal{O}\left(\frac{\rho^2\|R_{12}\|^2}{[1-(\rho')^2][\inf(L_{11})]^2}\right) \quad (6) \\
&\leq \sigma_j(R_{11})^{-1}\left(\frac{\|R_{22}\|}{\inf(R_{11})}\right)^2\mathcal{O}\left(\frac{\|R_{12}\|^2}{[1-(\rho')^2][\inf(L_{11})]^2}\right) \\
&\leq \sigma_j(R_{11})^{-1}\left(\frac{\sigma_{k+1}}{\sigma_k}\right)^2\mathcal{O}\left(\frac{n^{\frac{5}{2}}\|R_{12}\|^2}{[1-(\rho')^2][\inf(L_{11})]^2}\right). \quad (7)
\end{aligned}$$

The inequalities (5) and (6) follow from Theorem 2.1 of [13], and (7) follows from the bounds (2) and (3). □

In lieu of Theorem 3.1 of [13] in the proof above, we could have used a result of Eisenstat and Ipsen, Corollary 5.4 in [6]. It provides a slightly better bound, but the improvement would not show up in our proof since we take only a first order approximation in going to the big oh notation.

Theorem 2.1 shows that the convergence of $(\sigma_j(L_{11})^{-1} - \sigma_j^{-1})/\sigma_j^{-1}$ for $j = 1, \ldots, k$ are all quadratic in the gap ratio σ_{k+1}/σ_k. Hence if the gap ratio is small, then all of the singular values are likely well approximated. The theorem

does not address individual elements, only the upper and lower blocks. But asymptotic results describing how the norm of the matrix is being concentrated on the diagonal may be found in [4] and [11]. Most of the examples Stewart considered in [19] have at least one large gap in the singular values, and Theorem 2.1 indicates why the (truncated) pivoted QLP decomposition tends to do well in such cases. Low-rank problems are certainly one such case. Note that Theorem 2.1 is actually just a restatement of part of Theorem 3.3 of [11], which deals with the full pivoted QLP decomposition. This is because the L_{11} matrix is the same whether we truncate or not.

2.3 Operation Counts

In addition to being accurate, the truncated pivoted QLP algorithm is fast when k is small. To QR factorize k columns of an m-by-n matrix using Householder transformations requires about $2mnk - k^2(m+n) + 2/3k^3$ flam, where a flam is a floating-point addition combined with a floating-point multiplication (see p. 96 of [18]). Note that pivoting adds only an $\mathcal{O}(mn)$ term, so we need not distinguish when we are pivoting. Computing the complete QLP factorization requires computing all n columns of the QR factorization of the m-by-n matrix A and computing all n columns of the QR factorizatoin of the n-by-n matrix R^T. This gives a total flam count of

$$2mn^2 - n^2(m+n) + \frac{2}{3}n^3 + 2n^3 - n^2(n+n) + \frac{2}{3}n^3 = mn^2 + \frac{1}{3}n^3.$$

Computing the truncated QLP factorization, in which we compute only the first r rows of R, requires computing the first r columns of the QR factorization of the m-by-n matrix A and computing all r columns of the QR factorization of the n-by-r matrix $(R_{11}^T R_{12}^T)^T$). This gives a total flam count of

$$2mnr - r^2(m+n) + \frac{2}{3}r^3 + 2nr^2 - r^2(n+r) + \frac{2}{3}r^3 = 2mnr - mr^2 + \frac{1}{3}r^3.$$

If r is small ($r \ll n/2$), then the truncated QLP can be computed in $\mathcal{O}(mn)$ time compared to $\mathcal{O}(mn^2)$ time for the full QLP decomposition. This is a huge savings.

2.4 A Numerical Example

To show an example, we start with a 100-by-100 diagonal matrix having entries 100, 10, and the other ninety-eight evenly spaced between 10^{-2} and 10^{-8}. We then multiply by a random orthonormal matrix to obtain the matrix A we assume that we are initially given. Say we want good approximations to the large singular values, and that we do not care about the noise.

After taking the first two rows of the QR factorization of A, we see a minor gap in the first two R-values. We therefore halt this step and perform the next

step, computing the first two columns of L. The computed portions of R and L are given below. (The ellipsis [three dots] in each row of R stands for the ninety-eight other entries in that row.)

$$R = \begin{pmatrix} -94.4323 & 6.9336 & \cdots \\ & -8.8059 & \cdots \end{pmatrix} \qquad L = \begin{pmatrix} 99.9971 & 0.7519 \\ & 10.0002 \end{pmatrix}$$

From the first two L-values, we can see that the gap ratio is on the order of $1/10$ and that 99.9971 and 10.0002 are good approximations to the first two singular values. Perhaps we were expecting a smaller gap ratio to indicate the beginning of the noise. Let us compute an additional row of R and column of L. (Now each ellipsis in R stands for ninety-seven entries.)

$$R = \begin{pmatrix} -94.4323 & 6.9336 & 3.3312 & \cdots \\ & -8.8059 & 0.5307 & \cdots \\ & & 0.0984 & \cdots \end{pmatrix} \quad L = \begin{pmatrix} 99.9971 & 0.7519 & -0.0006 \\ & 10.0002 & -0.0009 \\ & & -0.0997 \end{pmatrix}$$

A gap ratio of about 10^{-3} between the second and third L-values is more like what we we expecting to separate the signal from the noise. We therefore take 99.9971 and 10.0002 as the two significant singular values of A.

Using the results from section 2.3, we see that the truncated pivoted QLP decomposition that we have computed (with $m = 10^2$, $n = 10^2$, and $r = 3$) requires about 6×10^4 flam. By contrast, the full pivoted QLP decomposition would have required about $4/3 \times 10^6$ flam. The truncated version is about twenty-two times faster.

References

[1] C. H. Bischof and C. Van Loan. The WY representation for products of householder matrices. *SIAM Journal on Scientific and Statistical Computing*, 8:s2–s13, 1987.

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

[3] S. Chandrasekaran and I.C.F. Ipsen. On rank-revealing factorisations. *SIAM Journal on Matrix Analysis and Applications*, 15(2):592–622, 1994.

[4] S. Chandrasekaran and I.C.F. Ipsen. Analysis of a QR algorithm for computing singular values. *SIAM Journal on Matrix Analysis and Applications*, 16(2):520–535, 1995.

[5] James W. Demmel. *Applied Numerical Linear Algebra.* SIAM, Philadelphia, 1997.

[6] S. Eisenstat and I. Ipsen. Relative perturbation techniques for singular value problems. *SIAM Journal on Numerical Analysis*, 32:1972–1988, 1995.

[7] L. V. Foster. Rank and null space calculations using matrix decomposition without column interchanges. *Linear Algebra and Its Applications*, 74:47–71, 1986.

[8] G. Golub and C. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, third edition, 1996.

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

[10] S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin. A theory of pseudoskeleton approximations. *Linear Algebra and Its Applications*, 261:1–21, 1997.

[11] D. A. Huckaby and T. F. Chan. On the convergence of Stewart's QLP algorithm for approximating the SVD. CAM report #02-29, University of California, Los Angeles, 2002.

[12] W. Kahan. Numerical linear algebra. *Canadian Math. Bulletin*, 9:757–801, 1966.

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

[14] R. S. Schreiber and C. Van Loan. A storage-efficient WY representation for products of householder transformations. *SIAM Journal on Scientific and Statistical Computing*, 10:52–57, 1989.

[15] G. W. Stewart. *Introduction to Matrix Computations.* Academic Press, New York, 1973.

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

[17] G. W. Stewart. Updating a rank-revealing ULV decomposition. *SIAM Journal on Matrix Analysis and Applications*, 14:494–499, 1993.

[18] G. W. Stewart. *Matrix Algorithms, Volume 1: Basic Decompositions.* SIAM, Philadelphia, 1998.

[19] G. W. Stewart. The QLP approximation to the singular value decomposition. *SIAM Journal on Scientific Computing*, 20(4):1336–1348, 1999.

[20] Lloyd N. Trefethen and David Bau, III. *Numerical Linear Algebra.* SIAM, Philadelphia, 1997.