TR-CS-98-14

A Completely Rank Revealing Quotient $URV$ Decomposition

Michael Stewart

December 1998

Joint Computer Science Technical Report Series
Department of Computer Science
Faculty of Engineering and Information Technology
Computer Sciences Laboratory
Research School of Information Sciences and Engineering
This technical report series is published jointly by the Department of Computer Science, Faculty of Engineering and Information Technology, and the Computer Sciences Laboratory, Research School of Information Sciences and Engineering, The Australian National University.

Please direct correspondence regarding this series to:

Technical Reports
Department of Computer Science
Faculty of Engineering and Information Technology
The Australian National University
Canberra ACT 0200
Australia

or send email to:

Technical.Reports@cs.anu.edu.au

A list of technical reports, including some abstracts and copies of some full reports may be found at:

http://cs.anu.edu.au/techreports/

Recent reports in this series:


A Completely Rank Revealing Quotient $URV$ Decomposition

Michael Stewart\textsuperscript{1}

September 1, 1998

This paper introduces a completely rank revealing complete orthogonal quotient decomposition for a pair of rectangular matrices. It reliably reveals an approximation of the minimum distance from a matrix pair with a prescribed quotient SVD structure. The approximation gives the true minimum distance up to a small constant factor that is independent of the sizes of the matrices. Consequently the decomposition is well suited to the recovery of the non-generic quotient SVD structure of a pair of matrices that have been corrupted by errors. The use of a completely rank revealing decomposition is shown to improve estimates of matrix range space intersections for a system identification problem.

1 Introduction

Let $A$ and $B$ be $m_a \times n$ and $m_b \times n$ matrices. Define

$$C = \begin{bmatrix} A \\ B \end{bmatrix}.$$  

If

$$r_a = \text{rank}(A), \quad r_b := \text{rank}(B), \quad r_c = \text{rank}(C)$$

then the dimension of the intersection of the row subspaces of $A$ and $B$ is

$$r_i = r_a + r_b - r_c.$$ \hspace{1cm} (1)

The quotient (or generalized) SVD (QSVD), [10, 12], is a decomposition that reveals these ranks. It has been defined as

$$U^T A Q = \begin{bmatrix} \Sigma_A R & 0 \\ r & n - r \end{bmatrix} \quad V^T B Q = \begin{bmatrix} \Sigma_B R & 0 \\ r & n - r \end{bmatrix}$$ \hspace{1cm} (2)

where

$$\Sigma_A = \begin{bmatrix} I_A & & \\
& S_A & \\
& & 0_A \end{bmatrix} \quad \Sigma_B = \begin{bmatrix} 0_B & & \\
& S_B & \\
& & I_B \end{bmatrix}$$

with square, diagonal, positive definite $S_A$ and $S_B$ typically chosen, by appropriate scaling on the rows of $R$, to satisfy $S_A^2 + S_B^2 = I$. The matrices $U$, $V$ and $Q$ are orthogonal and $R$ is $r_c \times r_c$, triangular and invertible.

\textsuperscript{1}Computer Sciences Laboratory, RSISE, Australian National University, Canberra ACT 0200, Australia, email: stewart@discus.anu.edu.au
The partitioning is such that \( S_A \) and \( S_B \) are the same size, \( r_i \). The identity matrices \( I_B \) and \( I_A \) are \((r_b - r_i) \times (r_b - r_i)\) and \((r_a - r_i) \times (r_a - r_i)\). The zero blocks \( 0_A \) and \( 0_B \) are \((m_a - r_a) \times (r_b - r_i)\) and \((m_b - r_b) \times (r_a - r_i)\) respectively. In addition to revealing the ranks of \( A \) and \( B \), the decomposition shows that \( r_i \) is the dimension of the intersection of the range spaces of \( A \) and \( B \). Generically, for \( m_a + m_b < n, r_i = 0 \) and \( r_a = m_a + m_b \).

Several of the algorithms proposed for computing the QSVD involve an orthogonal pre-processing step that computes orthogonal transformations \( U \), \( V \) and \( Q \) such that

\[
\begin{bmatrix}
U^T A Q \\
V^T B Q
\end{bmatrix}
\]

has a structure that either reveals the block sizes associated with (2) or simplifies the application of an iterative method for computing \( S_A \), \( S_B \) and \( R \).

If we limit ourselves to such orthogonal transformations by requiring that \( R = I \) then the most condensed form we can compute is

\[
\Sigma_A = \begin{bmatrix}
A_{11} & A_{12} & 0 \\
0 & A_{22} & 0 \\
0 & 0 & 0_A
\end{bmatrix} \quad \Sigma_B = \begin{bmatrix}
0_B & 0 & 0 \\
0 & B_{22} & 0 \\
B_{31} & B_{32} & B_{33}
\end{bmatrix}
\]  

(3)

where \( A_{11}, A_{22}, B_{22} \) and \( B_{33} \) are all square and upper triangular with full rank. A standard method for computing this starts with an orthogonal rank revealing compression of the columns of \( A \)

\[ AW = A \begin{bmatrix} W_1 & W_2 \end{bmatrix} = \begin{bmatrix} A_1 & 0 \end{bmatrix} \]

where \( A_1 \) has full rank. The block \( B_{33} \) is then obtained from a rank revealing factorization

\[
\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} B W_2 \begin{bmatrix} \hat{W}_1 & \hat{W}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ B_{33} & 0 \end{bmatrix}.
\]

Finally, \( B_{22} \) is determined by a similar rank revealing factorization of \( V_1^T B W_1 \) and a transformation, \( U \), can be introduced to give \( A_1 \) the required triangular structure.

The connection with (2) is not hard to see: the zero pattern and the full rank condition on the upper triangular blocks guarantee the existence of a nonorthogonal matrix formed from elementary elimination matrices which can be applied to \( \Sigma_A \) and \( \Sigma_B \) defined by (3) to zero \( A_{12} \), \( B_{31} \) and \( B_{32} \) without changing \( A_{11}, A_{22} \) and \( B_{22} \). Such a transformation has the form

\[
X = \begin{bmatrix}
I & 0 & 0 \\
0 & I & 0 \\
-B_{31} A_{11}^{-1} & -B_{31} A_{12} & B_{33}^{-1}
\end{bmatrix} \begin{bmatrix}
A_{11}^{-1} & -A_{11}^{-1} A_{12} & 0 \\
0 & I & 0 \\
0 & 0 & I
\end{bmatrix}
\]

giving

\[
\Sigma_A X = \begin{bmatrix}
I_A & 0 & 0 \\
0 & A_{22} & 0 \\
0 & 0 & 0_A
\end{bmatrix} \quad \Sigma_B X = \begin{bmatrix}
0_B & 0 & 0 \\
0 & B_{22} & 0 \\
0 & 0 & I_B
\end{bmatrix}.
\]

The matrices \( R \) and \( Q \) can be obtained from \( X = R Q \). Further transformations obtained from the SVD of \( B_{22}^{-1} A_{22} \) complete the diagonalization and it is easy to show that the singular
values of $B^{-1}_{22}A_{22}$ are equal to the diagonal elements of $S^{-1}_B S_A$. Clearly (3) is a generalized URV decomposition which gives all the rank information of (2).

This is a highly condensed derivation of the QSVD. A significant omission is that in the presence of errors on $A$ and $B$ we must allow for small nonzero elements in some of the zero blocks of (3) to reveal a nearby QSVD structure. How this should be done to reveal QSVD structure effectively is the topic of this paper.

In the constructive derivation of (3), we have seen that a quotient decomposition can be computed by starting with a rank revealing factorization of $A$. A decomposition computed in this way nominally reveals the same subspace information as the complete QSVD. However, sensitivity of the row subspace of $A$ can make the additional rank decisions required to compute (3) very difficult. We illustrate the difficulty with a small example.

**Example 1** Consider the matrix pair

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \delta & \epsilon & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where $0 < \epsilon < \delta < 1$. We assume that $\delta$ is significantly smaller than 1 but that it is large enough that $A$ can be considered to have full rank. We assume that $\epsilon$ is small enough that it is of the same order as the tolerance used in rank decisions. A perturbation to $A$ of norm $\epsilon$ clearly results in two full rank matrices with a one dimensional row subspace intersection.

Consider the orthogonal transformation given by the $LQ$ factorization of $A$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \delta & \epsilon & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{\delta}{\sqrt{\delta^2 + \epsilon^2}} & \frac{\epsilon}{\sqrt{\delta^2 + \epsilon^2}} & 0 & 0 \\ \frac{\epsilon}{\sqrt{\delta^2 + \epsilon^2}} & \frac{\delta}{\sqrt{\delta^2 + \epsilon^2}} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \delta & \epsilon \\ 0 & \sqrt{\delta^2 + \epsilon^2} & 0 & 0 \\ 0 & 0 & \frac{\delta}{\sqrt{\delta^2 + \epsilon^2}} & -\frac{\epsilon}{\sqrt{\delta^2 + \epsilon^2}} \\ 0 & 0 & 0 & 1 \end{bmatrix}. $$

Having computed a rank revealing factorization of $A$, we proceed by determining the rank of $BW_2$. In this case

$$BW_2 = \begin{bmatrix} \frac{-\epsilon}{\sqrt{\delta^2 + \epsilon^2}} & 0 \\ 0 & 1 \end{bmatrix}. $$

If $\epsilon = 16 - 1$ and $\delta = 1 - 8$ then

$$BW_2 = \begin{bmatrix} -1 & 8 \\ 0 & 1 \end{bmatrix}. $$

Without the perturbation, $\epsilon$, $BW_2$ will be rank deficient. However if $\epsilon$ is not zero and $\delta$ is at all small, we get a very hard rank decision. Since we have assumed that $\delta$ is greater than the tolerance used in rank decisions, we would conclude that $BW_2$ has full rank and that $r_2 = 4$. This implies that $r_1 = 0$, so that the algorithm completely misses the possibility that there is a nontrivial row subspace intersection achieved by matrices within $O(\epsilon)$ of $A$ and $B$. The end result is misleadingly partitioned decomposition that fails to reveal an interesting and potentially useful feature of $A$ and $B$. ■
The difficulty illustrated by this example is not a lack of numerical stability. In fact if \( n > m_a + m_b \), \( A \) and \( B \) have intersecting row subspaces only on a set of matrices that have measure zero in the space of all matrix pairs. Thus an arbitrarily small perturbation suffices to give a pair for which \( r_i = 0 \). The point is not that the algorithm is unstable in the sense of backward error; but that it tends to find a nearby QSVD structure that is less interesting and possibly less useful in some application than the structures exhibited by other nearby matrix pairs. Other algorithms can suffer from similar problems.

To give a basis for comparing methods for estimating QSVD structure in the presence of errors, we assume that \( A \) and \( B \) come from a model of the form

\[
A = \hat{A} + E, \quad B = \hat{B} + F
\]

where \( E \) and \( F \) are perturbations corrupting the matrices \( \hat{A} \) and \( \hat{B} \). Whatever the original structure of \( \hat{A} \) and \( \hat{B} \), almost all possible perturbations will result in \( A \) and \( B \) for which \( A \), \( B \) and \( C \) all have full rank. This is the generic QSVD structure of a matrix pair.

In attempting to recover the original, possibly non-generic structure of \( \hat{A} \) and \( \hat{B} \), we would like to be able to determine if small perturbations to \( A \) and \( B \) can be chosen to give a prescribed rank deficient QSVD structure. This is the standard by which we will measure the merits of algorithms for recovering QSVD structure from perturbed matrices: a completely rank revealing algorithm for computing a quotient decomposition should be able to determine reliably if a matrix pair is close to a matrix pair with a particular prescribed QSVD structure. Taking the ordinary singular value decomposition and the Eckart-Young theorem as a benchmark, we might even hope to compute a decomposition that reveals the smallest perturbations that give a particular structure.

A number of algorithms have been proposed for QSVD computation and for the related problem of finding range space intersections. One common first step is to start with a computation of the row subspace of \( A \). After fixing an orthonormal basis for this space and observing how it relates to \( B \), the best we can hope to do is to determine if a small perturbation to \( B \) alone gives a nonempty intersection with the fixed row subspace of \( A \). If the original basis for the row subspace of \( A \) was at all ill-conditioned, small perturbations to \( A \) can change the space significantly. Considering only the possibility of perturbations on \( B \) can give misleading results.

Another approach is to find orthonormal bases for the row subspaces of both \( A \) and \( B \) and then use the principal angles between the two subspaces to estimate the intersection dimension, [4]; this suffers from a similar problem when either of the subspaces are sensitive. In some cases it is also possible to compute a complete QSVD without any initial rank decisions and then try to determine directly the number of nontrivial quotient singular values of \( A \) and \( B \). Quotient singular values can be sensitive to perturbations, [8], and when using this method it can be difficult to relate their magnitudes to the sizes of \( E \) and \( F \). This difficulty is particular worrisome when \( m_a + m_b \leq n \); under these circumstances there is generically no intersection and the generalized singular values are generically all either 1 or 0. None of these methods can reliably decide if two matrices are consistent with a model of the form (5) for particular \( r_a \), \( r_b \) and \( r_i \). Further, even if an error tolerance is chosen so that these methods succeed in estimating the model ranks correctly, there is no guarantee that the estimated \( \hat{A} \) and \( \hat{B} \) are as close to the original data as possible; this can result in
loss of accuracy in the subspace estimates.

Fortunately, the existence of a method that can reliably recover rank deficient QSVD structure of nearby matrix pairs is a trivial consequence of (1). Given $A$ and $B$ satisfying (5), each of the ranks can be determined using completely independent rank decisions. Assume that we are given a prescribed set of ranks $r_a$, $r_b$ and $r_c$ and an error tolerance. We wish to find small perturbations to $A$ and $B$ that give the desired non-generic QSVD structure. Using an ordinary rank revealing factorization such as the SVD we can compute

$$CU = \begin{bmatrix} A \\ B \end{bmatrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} A_1 & E_2 \\ B_1 & F_2 \end{bmatrix}$$

where

$$\left\| \begin{bmatrix} E_2 \\ F_2 \end{bmatrix} \right\|_2 = \sigma_{r_c+1}(C),$$

$U^TU = I$ and $A_1$ and $B_1$ are $m_a \times r_c$ and $m_b \times r_c$ respectively. By setting the $E$ and $F$ blocks to zero, we get a minimal perturbation to induce the desired degree of rank deficiency in $C$. If we choose $E_1$ and $F_1$ so that

$$\text{rank}(A_1 - E_1) = r_a, \quad \text{rank}(B_1 - F_1) = r_b$$

then

$$C - \begin{bmatrix} E_1 & E_2 \\ F_1 & F_2 \end{bmatrix} U^T = \begin{bmatrix} A_1 - E_1 & 0 \\ B_1 - F_1 & 0 \end{bmatrix} U^T$$

has at least the desired degree of rank deficiency. (After the transformation by $U$, the blocks $A_1$ and $B_1$ could have a rank deficiency that is higher than desired; however, if this is the case then an arbitrarily small perturbation gives a pair for which $A$ and $B$ have the desired rank.)

For any $E$ and $F$, if $A + E$ and $B + F$ are a pair with ranks $r_a$, $r_b$ and $r_c$ then

$$\left\| \begin{bmatrix} E \\ F \end{bmatrix} \right\|_2 \geq \max(\sigma_{r_c+1}(C), \sigma_{r_a+1}(A), \sigma_{r_b+1}(B))$$

where if $r_a = \min(m_a, n)$ or $r_b = \min(m_b, n)$ we take $\sigma_{r_a+1}(A) = 0$ or $\sigma_{r_b+1}(B) = 0$. It follows that

$$\left\| \begin{bmatrix} E_1 & E_2 \\ F_1 & F_2 \end{bmatrix} \right\|_2 \leq \sigma_{r_c+1}(C) + \sigma_{r_a+1}(A) + \sigma_{r_b+1}(B) \leq 3 \left\| \begin{bmatrix} E \\ F \end{bmatrix} \right\|_2.$$  \hspace{1cm} (6)

Thus, within a factor of 3, an algorithm based on rank decisions on $C$, $A$ and $B$ can be used to find a matrix pair which is as close as possible among those with a given QSVD structure. Given an estimate of the size of the errors corrupting $A$ and $B$, we can reliably show when two matrices are consistent with (5) for given ranks $r_a$, $r_b$ and $r_c$.

This observation is the main contribution of this paper. We will extend these ideas and develop a completely rank revealing quotient decomposition of the form (3) in §2. While we do not develop an updating algorithm, the decomposition might be considered a quotient generalization of a URV decomposition. It can also be viewed as a complete orthogonal
version of a generalized QR factorization as in [9]. Related generalizations have been proposed in [3, 5, 6, 11]. The algorithm described in [1] and implemented in LAPACK as a preprocessing step for a full QSVD is essentially the first part of a quotient generalization of a complete orthogonal decomposition. While less contrived examples are not as striking as Example 1, in §3 we will show that the decomposition can give estimates range space intersections that are superior to each of the alternate algorithms mentioned in this section.

2 The Decomposition

Without loss of generality, we assume that \( n \geq m_a, m_b \). If this condition fails, then \( A \) and \( B \) can be replaced by square matrices through the use of the QR factorization. As we have noted, a matrix pair has QSVD structure (2) if and only if

\[
\text{rank}(A) = r_a, \quad \text{rank}(B) = r_b, \quad r_a + r_b - r_i = \text{rank}(C).
\]

Given these prescribed ranks, the following algorithm computes a decomposition that comes within a constant factor of finding the smallest perturbations that give the appropriate QSVD structure.

**Algorithm 1** Given two \( m_a \times n \) and \( m_b \times n \) matrices \( A \) and \( B \) with \( n \geq m_a, m_b \) and ranks \( r_a, r_b, \) and \( r_c, \) the following algorithm computes a completely rank revealing quotient decomposition.

1. Use the SVD to perform an orthogonal rank revealing compression on the columns of \( C \)

\[
\begin{bmatrix} A \\ B \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} = \begin{bmatrix} C_1 & F_1 \end{bmatrix}
\]

where \( C_1 \) and \( F_1 \) are \( (m_a + m_b) \times r_c \) and \( (m_a + m_b) \times (n - r_c) \) respectively with

\[
\|F_1\| \leq \sigma_{r_a+1}(C).
\]

Let

\[
\begin{bmatrix} A \\ B \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} = \begin{bmatrix} A^{(1)} & E^{(1)} \\ B^{(1)} & F^{(1)} \end{bmatrix}
\]

where \( A^{(1)} \) is \( m_a \times r_c \) and \( B^{(1)} \) is \( m_b \times r_c \).

2. Perform an orthogonal rank revealing triangular decomposition on \( A \) to get

\[
\begin{bmatrix} A^{(2)}_{11} & A^{(2)}_{12} & E^{(2)}_{13} \\ 0 & E^{(2)}_{22} & E^{(2)}_{23} \\ B^{(2)}_{11} & B^{(2)}_{12} & F^{(2)}_{22} \end{bmatrix}
\]

where \( A^{(2)}_{11} \) is \( r_a \times r_a \). For the purpose of proving a distance bound we require that

\[
\|E^{(2)}_{22}\|_2 \leq \sigma_{r_a+1}(A^{(1)}) \leq \sigma_{r_a+1}(A).
\]
3. Compute a rank revealing factorization of

\[
\begin{bmatrix}
B_{11}^{(2)} & B_{12}^{(2)}
\end{bmatrix}
\]

applying the same right transformations to \( A \) to get

\[
\begin{bmatrix}
A_{11}^{(3)} & A_{12}^{(3)} & E_{13}^{(3)} \\
F_{21}^{(3)} & F_{22}^{(3)} & E_{23}^{(3)} \\
F_{11}^{(3)} & 0 & F_{13}^{(3)} \\
B_{21}^{(3)} & B_{22}^{(3)} & F_{23}^{(3)}
\end{bmatrix}
\]

where \( B_{22}^{(3)} \) is \( r_b \times r_b \) and

\[
\|F_{11}^{(3)}\|_2 \leq \sigma_{r_b+1}( \begin{bmatrix} B_{11}^{(2)} & B_{12}^{(2)} \end{bmatrix} ) \leq \sigma_{r_b+1}(B).
\]

4. Apply an orthogonal transformation from the right to get

\[
\begin{bmatrix}
A_{11}^{(4)} & 0 & E_{13}^{(4)} \\
E_{21}^{(4)} & E_{22}^{(4)} & E_{23}^{(4)} \\
F_{11}^{(4)} & F_{12}^{(4)} & F_{13}^{(4)} \\
B_{21}^{(4)} & B_{22}^{(4)} & F_{23}^{(4)}
\end{bmatrix}
\]

where \( A_{11}^{(4)} \) is \( r_a \times r_a \) and \( B_{22}^{(4)} \) is \( r_b \times (r_b - r_i) \).

5. Apply orthogonal transformations from the left to \( B \) to get

\[
\begin{bmatrix}
A_{11}^{(5)} & 0 & E_{13}^{(5)} \\
E_{21}^{(5)} & E_{22}^{(5)} & E_{23}^{(5)} \\
F_{11}^{(5)} & F_{12}^{(5)} & F_{13}^{(5)} \\
B_{21}^{(5)} & 0 & F_{23}^{(5)} \\
B_{31}^{(5)} & B_{32}^{(5)} & F_{33}^{(5)}
\end{bmatrix}
\]

where \( B_{22}^{(5)} \) is \( (r_b - r_i) \times (r_b - r_i) \) and \( B_{21}^{(5)} \) is \( r_i \times r_a \).

6. Apply orthogonal transformations from the right to introduce zeros into \( B_{21}^{(5)} \)

\[
\begin{bmatrix}
A_{11}^{(6)} & A_{12}^{(6)} & 0 & E_{11}^{(6)} \\
E_{21}^{(6)} & E_{22}^{(6)} & E_{23}^{(6)} & E_{24}^{(6)} \\
F_{11}^{(6)} & F_{12}^{(6)} & F_{13}^{(6)} & F_{14}^{(6)} \\
0 & B_{22}^{(6)} & 0 & F_{24}^{(6)} \\
B_{31}^{(6)} & B_{32}^{(6)} & B_{33}^{(6)} & F_{34}^{(6)}
\end{bmatrix}
\]

where \( B_{22}^{(6)} \) is \( r_i \times r_i \).
7. Finally perform a $QR$ factorization on the $r_a \times r_a$ block

$$
\begin{bmatrix}
A_{11}^{(6)} & A_{12}^{(6)}
\end{bmatrix}
$$

to get

$$
\begin{bmatrix}
A_{11}^{(7)} & A_{12}^{(7)} & 0 & E_{14}^{(7)} \\
0 & A_{22}^{(7)} & 0 & E_{24}^{(7)} \\
E_{31}^{(7)} & E_{32}^{(7)} & E_{33}^{(7)} & E_{34}^{(7)} \\
F_{11}^{(7)} & F_{12}^{(7)} & F_{13}^{(7)} & F_{14}^{(7)} \\
0 & B_{22}^{(7)} & 0 & F_{24}^{(7)} \\
B_{31}^{(7)} & B_{32}^{(7)} & B_{33}^{(7)} & F_{34}^{(7)}
\end{bmatrix}
$$

where $A_{22}^{(7)}$ is $r_i \times r_i$. The pair of matrices with the desired rank structure may be found by setting all $E$ and $F$ blocks to zero. ■

Step 1 ensures that

$$
\begin{bmatrix}
E_{14}^{(7)} \\
E_{24}^{(7)} \\
E_{34}^{(7)} \\
F_{14}^{(7)} \\
F_{24}^{(7)} \\
F_{34}^{(7)}
\end{bmatrix} \leq \sigma_{r_a+1}(C),
$$

Steps 2 and 3 give

$$
\begin{bmatrix}
E_{31}^{(7)} & E_{32}^{(7)} & E_{33}^{(7)}
\end{bmatrix} \leq \sigma_{r_a+1}(A)
$$

and

$$
\begin{bmatrix}
F_{11}^{(7)} & F_{12}^{(7)} & F_{13}^{(7)}
\end{bmatrix} \leq \sigma_{r_b+1}(B).
$$

If $E$ and $F$ are perturbations such that $A + E$ and $B + F$ have the desired QSVD structure then (6) implies that

$$
\begin{bmatrix}
0 & 0 & 0 & E_{14}^{(7)} \\
0 & 0 & 0 & E_{24}^{(7)} \\
E_{31}^{(7)} & E_{32}^{(7)} & E_{33}^{(7)} & E_{34}^{(7)} \\
F_{11}^{(7)} & F_{12}^{(7)} & F_{13}^{(7)} & F_{14}^{(7)} \\
0 & 0 & 0 & F_{24}^{(7)} \\
0 & 0 & 0 & F_{34}^{(7)}
\end{bmatrix} \leq \sigma_{r_a+1}(C) + \sigma_{r_a+1}(A) + \sigma_{r_b+1}(B) \leq 3 \begin{bmatrix} E \\ F \end{bmatrix} \, _2.
$$

Thus, the algorithm finds a pair of matrices with a particular QSVD structure that is nearly as close as possible to $A$ and $B$.

In describing the algorithm we have assumed that

$$
r_a \leq \min(m_a, r_c), \quad r_b \leq \min(m_b, r_c)
$$

8
and 
\[ r_c \leq \min(n, m_a + m_b) \]
so that each of the rank decisions is nontrivial. If any of these conditions fails to hold, then it is not hard to simplify the algorithm by eliminating one or more of the rank decisions.

The implementation of the first three steps of the algorithm deserves special comment. In addition to the transformation \( Q \), step 1 can make use of transformations from the left that do not mix the rows of \( A \) and \( B \). In fact, to determine the numerical rank of \( C \) more general two sided transformations will be necessary. In practice, it might be necessary to compute a complete rank revealing decomposition separately and use the computed null space to give the transformation in step 1. Steps 2 and 3 can make use of in-place rank revealing decompositions. The SVD can be used and will guarantee the bounds we have claimed. Other methods such as pivoted \( QR \) can also be used to get a typically reliable algorithm. In this case, the quality of the bound that is analogous to (7) would depend on the ability of the pivoted \( QR \) factorization to detect rank deficiency. Under most circumstances, the use of a column-pivoted \( QR \) factorization is likely to work reasonably well.

There is nothing in this formulation to prevent rank deficiency or near rank deficiency in \( A_{11}, A_{22}, B_{22}, \) or \( B_{33} \). However ill-conditioning in \( A_{11}, A_{22}, \) or \( B_{22} \) indicates that small perturbations will decrease the rank of \( A \) or \( B \) even further. Ill-conditioning in \( B_{33} \) indicates that a small perturbation to \( B \) decreases the rank of \( C \). The algorithm can spot these situations in determining appropriate numerical ranks for \( A, B \) and \( C \). If the rank decisions are clear cut, then we can expect \( A_{11}, A_{22}, B_{22} \) and \( B_{33} \) to be not only nonsingular but also well conditioned. Thus we can expect to find a QSVD structure with minimal ranks for \( A, B \) and \( C \). This is equivalent to finding the largest intersection among the set of most rank deficient nearby matrix pairs.

3 An Application

We consider an application in which the intersection of the row subspaces of two matrices can be used to identify a state space model from observed input/output vectors. Consider the state space model
\[ x_{k+1} = Ax_k + Bu_k \]
\[ y_k = Cx_k + Du_k, \] (8)
where \( x_k \) is \( r \times 1 \), \( u_k \) is \( m \times 1 \) and \( y_k \) is \( p \times 1 \). Assuming we have observations of the input and output vectors, \( u_k \) and \( y_k \), the identification problem is to find an order, \( r \), and system matrices \( A, B, C \) and \( D \) that satisfy, or approximately satisfy, (8) for some sequence of \( r \times 1 \) state vectors \( x_k \).

The quotient decomposition applies naturally to a system identification algorithm developed in [7]. The approach can be characterized by two steps: find an estimate of the state
sequence $x_k$, and then obtain the system matrices from the least squares problem

$$
\begin{bmatrix}
  x_{k+i+j-1} & \cdots & x_{k+i+1} \\
  y_{k+i+j-2} & \cdots & y_{k+i}
\end{bmatrix}
= \begin{bmatrix}
  A & B \\
  C & D
\end{bmatrix}
\begin{bmatrix}
  x_{k+i+j-2} & \cdots & x_{k+i} \\
  y_{k+i+j-2} & \cdots & y_{k+i}
\end{bmatrix}.
$$

(9)

The index $k$ is the time at which observations begin and $k + i + j - 1$ is the time at which the latest observations have been made.

Define the $m \times j$ block Toeplitz matrices

$$
U_k = \begin{bmatrix}
  u_{k+j-1} & u_{k+j-2} & \cdots & u_k \\
  u_{k+j} & u_{k+j-1} & \cdots & u_{k+1} \\
  \vdots & \vdots & \ddots & \vdots \\
  u_{k+j+i-2} & u_{k+j+i-3} & \cdots & u_{k+i-1}
\end{bmatrix}
$$

and

$$
Y_k = \begin{bmatrix}
  y_{k+j-1} & y_{k+j-2} & \cdots & y_k \\
  y_{k+j} & y_{k+j-1} & \cdots & y_{k+1} \\
  \vdots & \vdots & \ddots & \vdots \\
  y_{k+j+i-2} & y_{k+j+i-3} & \cdots & y_{k+i-1}
\end{bmatrix}
$$

and

$$
T_k = \begin{bmatrix}
  U_k \\
  Y_k
\end{bmatrix}.
$$

The following theorem from [7] provides a means for generating an appropriate sequence of state vectors.

**Theorem 1** Let the vectors $u_k$ and $y_k$ be generated by

$$
x_{k+1} = Ax_k + Bu_k
$$

$$
y_k = C x_k + Du_k
$$

where the rank of

$$
\begin{bmatrix}
  C^T & A^T C^T & \cdots & (A^T)^{r-1} C^T
\end{bmatrix}
$$

is $r$.

Let

$$
X_k = \begin{bmatrix}
  x_{k+j-1} & x_{k+j-2} & \cdots & x_k
\end{bmatrix}
$$

and

$$
X_{k+i} = \begin{bmatrix}
  x_{k+i+j-1} & x_{k+i+j-2} & \cdots & x_{k+i}
\end{bmatrix}.
$$

For $i \geq r$, if $\text{rank}(X_k) = \text{rank}(X_{k+i}) = r$ and the matrices

$$
\begin{bmatrix}
  U_k \\
  X_k
\end{bmatrix}, \quad \begin{bmatrix}
  U_{k+i} \\
  X_{k+i}
\end{bmatrix}, \quad \begin{bmatrix}
  U_k \\
  U_{k+i}
\end{bmatrix}
$$

(11)
all have full rank $mi + r$, $mi + r$ and $2mi + r$ respectively then $T_k$ and $T_{k+i}$ both have rank $mi + r$ and the intersection of the span of the rows of $T_k$ and $T_{k+i}$ has dimension $r$. Further there is a basis, $X$, of the intersection for which

$$X = \begin{bmatrix} x_{k+i-1} & x_{k+i-2} & \cdots & x_{k+i} \end{bmatrix}$$

and different bases for this space correspond to state vector sequences of models with equivalent input/output behavior under a transformation of the form

$$\{SAS^{-1}, SB, CS^{-1}, D\}.$$ 

The rank conditions hold generically for a persistently exciting input. More details may be found in [7].

To compare the decomposition to other methods for extracting a row subspace intersection we consider the system defined by

$$A = \begin{bmatrix} .4 & 0 & .8 \\ .4 & .4 & -.4 \\ .4 & 0 & .4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ -4 & 2 \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & -1 & 0 \\ 1 & -2 & -1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$ 

The system can be verified to be stable with its largest eigenvalue having magnitude strictly less than 1. The observability condition (10) is also easy to verify.

We generated a sequenced of input vectors $u_k$ with elements that were randomly generated according to a zero mean normal distribution with variance 1. The initial state state vector was chosen as $x_1 = 0$. A sequence of output vectors, $y_k$, were generated by (8). Before applying the identification algorithm, each component of the input and output vectors were perturbed by zero mean unit variance normal noise scaled by factor of .1. We formed Toeplitz matrices $T_k$ and $T_{k+i}$ with $j = 100$ and $i = 10$. This resulted in Toeplitz matrices with a noise component roughly one order of magnitude below the smallest signal singular values of $T_k$ or $T_{k+i}$.

Over the course of 100 runs for different randomly generated inputs, we compared the performance of four different methods for estimating a state sequence. The first method used the algorithm of this paper to directly estimate a row subspace intersection for the matrices $T_k$ and $T_{k+10}$. The second method used an intersection estimate proposed in [7]. This estimate has the interesting feature of leading to a compact variation of the least squares problem (9) in terms of the just the singular values and left singular vectors of the combined matrix

$$T = \begin{bmatrix} T_k \\ T_{k+10} \end{bmatrix}$$

and $T_k$; the intersection does not appear explicitly, but it can be generated for comparison purposes. Since the method can be formulated to avoid explicit use of the right singular vectors, it is not hard to prove that it is consistent, [2]. The third approach used a complete orthogonal quotient decomposition based on an initial rank decision on $T_k$. This is essentially
<table>
<thead>
<tr>
<th>Method</th>
<th>$|P(X) - P(\hat{X})|_2$</th>
<th>$|Y - \hat{Y}|_2 / |Y|_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0581</td>
<td>0.0356</td>
</tr>
<tr>
<td>2</td>
<td>0.0901</td>
<td>0.0360</td>
</tr>
<tr>
<td>3</td>
<td>0.1041</td>
<td>0.0429</td>
</tr>
<tr>
<td>4</td>
<td>0.0901</td>
<td>0.0359</td>
</tr>
</tbody>
</table>

Table 1: Errors for Different Intersection Estimates

the method that was shown to fail so dramatically in Example 1. Finally, the fourth approach was based on an algorithm proposed in [4]. We computed the principal angles between the subspaces and used the smallest three angles together with the corresponding subspace of the row space of $T_k$ to get an intersection estimate.

Given the known exact state sequence $X$, we measured the quantity

$$\sqrt{\frac{1}{100} \sum_{p=1}^{100} \| P(X) - P(\hat{X}_p) \|^2_2}$$

where $P(X)$ is the projection matrix for the row subspace of $X$ and $P(\hat{X}_p)$ is the projection for the row subspace of the estimated state subspace in run $p$.

In a similar manner, we averaged the squares of the output error residuals. That is, we formed a matrix

$$\hat{Y} = \begin{bmatrix} \hat{y}_k & \hat{y}_{k+1} \cdots \hat{y}_{k+j} \end{bmatrix}$$

by driving the system (8) for the estimated $A$, $B$, $C$ and $D$ with the original unperturbed input $u_k$. Forming a matrix $Y$ of exact outputs in a similar manner we computed the average relative output error

$$\sqrt{\frac{1}{100} \sum_{p=1}^{100} \left\| Y - \hat{Y}_p \right\|_2^2 / \left\| Y \right\|_2^2}.$$

These averaged results are shown in Table 3 for each of the four proposed methods listed in order.

The rough proportions between the averaged values in the table did not change significantly for separate experiments with 100 runs. There are several interesting features of the data. The most obvious is that the method of this paper gives significantly better subspace estimates. Improving subspace estimates by introducing a decomposition that is more consistent with the model (5) was a main goal of this paper. In this application, the decomposition appears to be successful.

However, it is interesting to note that the improved subspace estimates did not lead to a significant improvement in the output error residuals as compared with the method of [7]. It seems likely that an error analysis of the subspace identification algorithm would be required to explain this phenomenon. While the results are quite surprising, further attempts at explanation are almost certainly beyond the scope of this paper.

Finally, we note that method 3, in which a quotient decomposition is computed by starting with a rank and range space estimate of $A$, achieved the worst results. While it is
not obvious from the table, the worst-case results using this method were much worse than for the other three methods. In most of the runs it was competitive with the other methods. Much of the increased average error was due to the relatively few runs in which it achieved extremely poor results.

References


