THE AUSTRALIAN NATIONAL UNIVERSITY

TR-CS-98-13

Finding Near Rank Deficiency in Matrix Products

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:

TR-CS-98-12 Vadim Olshevsky and Michael Stewart. *Stable factorization of Hankel and Hankel-like matrices*. December 1998.

TR-CS-98-11 Michael Stewart. *An error analysis of a unitary Hessenberg QR algorithm*. December 1998.

TR-CS-98-10 Peter Strazdins. *Optimal load balancing techniques for block-cyclic decompositions for matrix factorization*. September 1998.

TR-CS-98-09 Jim Grundy, Martin Schwenke, and Trevor Vickers (editors). *International Refinement Workshop & Formal Methods Pacific '98 — Work-in-progress papers of IRW/FMP'98, 29 September – 2 October 1998, Canberra, Australia*. September 1998.

TR-CS-98-08 Jim Grundy and Malcolm Newey (editors). *Theorem Proving in Higher Order Logics: Emerging Trends — Proceedings of the 11th International Conference, TPHOLs'98, Canberra, Australia, September – October 1998, Supplementary Proceedings*. September 1998.

TR-CS-98-07 Peter Strazdins. *A comparison of lookahead and algorithmic blocking techniques for parallel matrix factorization*. July 1998.

Finding Near Rank Deficiency in Matrix Products[1]

Michael Stewart

December 1, 1998

This paper gives a theorem characterizing approximately minimal norm rank one perturbations E and F that make the product $(A+E)(B+F)^{\mathrm{T}}$ rank deficient. The theorem is stated in terms of the smallest singular value of a particular matrix chosen from a parameterized family of matrices by solving a nonlinear equation. Consequently, it is analogous to the special case of the Eckhart-Young theorem describing the minimal perturbation that induces an order one rank deficiency. While the theorem does not naturally extend to higher order rank deficiencies, it can be used to compute a complete orthogonal product decomposition to give improved practical reliability in revealing the numerical rank of AB^{T}.

1 Introduction

We assume that the $m_a \times n$ and $m_b \times n$ matrices A and B with $n > m_a, m_b$ come from a model of the form

$$A = \hat{A} + E, \qquad B = \hat{B} + F \tag{1}$$

where $\hat{A}$, $\hat{B}$ or $\hat{A}\hat{B}^{\mathrm{T}}$ exhibit some degree of rank deficiency and E and F are perturbations corrupting the exact data in $\hat{A}$ and $\hat{B}$. Without any loss of generality we assume that $m_a \leq m_b$. Thus A, B, and AB^{T} will generically have full rank even if rank deficiency of the underlying model implies that one or more of these perturbed matrices will be ill-conditioned.

Our primary goal is to recover the rank of $\hat{A}\hat{B}^{\mathrm{T}}$ given A and B and to find estimates of the corresponding range and null spaces. To give full generality, we might also be concerned with simultaneously finding an estimate of the ranks of $\hat{A}$ and $\hat{B}$.

Before moving on to the central and most difficult problem of estimating the product rank, we describe a complete orthogonal product decomposition to jointly reveal the ranks of $\hat{A}$, $\hat{B}$ and $\hat{A}\hat{B}^{\mathrm{T}}$ when the unperturbed matrices are available. The decomposition takes the form

$$U^{\mathrm{T}}\hat{A}Q = \begin{bmatrix} A_{11} & 0 & 0 & 0 & 0 \\ A_{21} & A_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad V^{\mathrm{T}}\hat{B}Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & B_{23} & 0 & 0 \\ 0 & B_{32} & B_{33} & B_{34} & 0 \end{bmatrix} \tag{2}$$

where U, V and Q are square and orthogonal, the columns of the matrices are partitioned in the same way and A_{11}, A_{22}, B_{32} and B_{23} are square and have full rank. If

$$r_a = \mathrm{rank}(\hat{A}), \qquad r_b = \mathrm{rank}(\hat{B}), \qquad r_p = \mathrm{rank}(\hat{A}\hat{B}^{\mathrm{T}})$$

[1]Computer Sciences Laboratory, RSISE, Australian National University, Canberra ACT 0200, Australia, email: stewart@discus.anu.edu.au

then A_{11} is $(r_a - r_p) \times (r_a - r_p)$, A_{22} and B_{32} are $r_p \times r_p$, B_{23} is $(r_b - r_p) \times (r_b - r_p)$ and B_{24} is $r_p \times r_p$.

One possible algorithm for computing this decomposition starts with an orthogonal rank revealing factorization of $\hat{A}$ to get

$$\begin{bmatrix} U^{(1)} & 0 \\ 0 & V^{(1)} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \hat{A} \\ \hline \hat{B} \end{bmatrix} Q^{(1)} = \left[\begin{array}{cc} A_1^{(1)} & 0 \\ 0 & 0 \\ \hline B_1^{(1)} & B_2^{(1)} \end{array}\right] \tag{3}$$

where $A_1^{(1)}$ is $r_a \times r_a$ and has full rank. Since

$$\operatorname{rank}\left(\hat{A}\hat{B}^{\mathrm{T}}\right) = \operatorname{rank}\left(A_1^{(1)}\left(B_1^{(1)}\right)^{\mathrm{T}}\right) = r_p$$

a rank revealing factorization of $B_1^{(1)}$ gives

$$\begin{bmatrix} I & 0 \\ 0 & V^{(2)} \end{bmatrix}^{\mathrm{T}} \left[\begin{array}{cc} A_1^{(1)} & 0 \\ 0 & 0 \\ \hline B_1^{(1)} & B_2^{(1)} \end{array}\right] Q^{(2)} = \left[\begin{array}{ccc} A_1^{(2)} & A_2^{(2)} & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & B_{13}^{(2)} \\ 0 & B_{22}^{(2)} & B_{23}^{(2)} \end{array}\right]$$

where $B_{22}^{(2)}$ is $r_p \times r_p$ and has full rank. Clearly $\operatorname{rank}(B_{13}^{(2)}) = r_b - r_p$. A rank revealing factorization of $B_{13}^{(2)}$ gives

$$\begin{bmatrix} I & 0 \\ 0 & V^{(3)} \end{bmatrix}^{\mathrm{T}} \left[\begin{array}{ccc} A_1^{(2)} & A_2^{(2)} & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & B_{13}^{(2)} \\ 0 & B_{32}^{(2)} & B_{23}^{(2)} \end{array}\right] Q^{(3)} = \left[\begin{array}{cccc} A_1^{(3)} & A_2^{(3)} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & B_{23}^{(3)} & 0 \\ 0 & B_{32}^{(3)} & B_{33}^{(3)} & B_{34}^{(3)} \end{array}\right]$$

where $B_{23}^{(3)}$ is $(r_b - r_p) \times (r_b - r_p)$ and $B_{32}^{(3)} = B_{22}^{(2)}$. A further transformation $U^{(3)}$ can be used to give $\hat{A}$ the desired block triangular structure. To get (2) the $r_p \times (n - r_a - r_b + r_p)$ matrix $B_{34}^{(3)}$ can be compressed into r_p nonzero columns using a further transformation $Q^{(4)}$. Clearly the singular values of the product $A_{22}B_{32}^{\mathrm{T}}$ are the nonzero singular values of $\hat{A}\hat{B}^{\mathrm{T}}$. Related decompositions may be found in [3, 1].

The above discussion shows that r_p can be found from the SVD of $B_1^{(1)}$. It can also be found directly from the singular values of $\hat{A}\hat{B}^{\mathrm{T}}$. However, without access to the unperturbed $\hat{A}$ and $\hat{B}$ neither of these approaches is entirely satisfactory. For the latter this can be seen from the two examples $A = B = \sqrt{\epsilon}$ and $A = 1$, $B = \epsilon$. The product singular value, ϵ, is the same, but in the second example A and B come from a model of the form (1) with $E, F = O(\epsilon)$ and $\operatorname{rank}(\hat{A}\hat{B}^{\mathrm{T}}) = 0$; in the first example they do not. Looking for small singular values in $B_1^{(1)}$ can also fail. The problem is that $B_1^{(1)}$ depends on a possibly sensitive estimate of the row subspace of A. We illustrate this with a small example.

Example 1 Consider the perturbed matrix pair

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \delta & \epsilon & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{4}$$

where $0 < \epsilon < \delta < 1$. The element ϵ represents a perturbation to

$$\hat{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \delta & 0 & 0 \end{bmatrix}.$$

The matrix $\hat{B}$ is unperturbed. We suppose that δ is significantly smaller than 1 but that it is large enough that A can be considered to have full rank. We assume that ϵ is small enough that it is of the same order as the tolerance used in rank decisions.

We consider the algorithm outlined in the derivation of (2) applied to the perturbed matrices A and B. The algorithm starts with an orthogonal transformation determined by the LQ factorization of A

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

This is the factorization step represented in (3). Having computed a trivial rank revealing factorization of A, we proceed by determining the rank of $B_1^{(1)}$. If $\epsilon = 1e-16$ and $\delta = 1e-8$ then

$$B_1^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{\epsilon}{\sqrt{\delta^2+\epsilon^2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1e-8 \end{bmatrix}.$$

Without the perturbation ϵ, $B_1^{(1)}$ will be exactly rank deficient. However if ϵ is not zero and δ is at all small, we get a very hard rank decision. Since we have assumed that δ is greater than the tolerance, we would conclude for these chosen values of ϵ and δ that $B_1^{(1)}$ has full rank and that $r_p = 2$. The end result is misleadingly partitioned decomposition that fails to reveal near rank deficiency in AB^{T}. ■

As an alternative to looking at the product singular values or at the singular values of $B_1^{(1)}$, we propose a generalization of the Eckhart-Young theorem to find nearly minimal perturbations E and F that make $(A-E)(B-F)^{\mathrm{T}}$ rank deficient. The ordinary Eckhart-Young theorem may be stated as follows. Let

$$A = U \begin{bmatrix} \Sigma & 0_{m_a, n-m_a} \end{bmatrix} Q^{\mathrm{T}}$$

with $U^{\mathrm{T}}U = I_{m_a}$, $Q^{\mathrm{T}}Q = I_n$ and

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

for $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{m_a} \geq 0$. If

$$\hat{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p, 0, 0, \ldots, 0)$$

and

$$\hat{A} = U \begin{bmatrix} \hat{\Sigma} & 0 \end{bmatrix} V^{\mathrm{T}}$$

then the Eckhart-Young theorem states that

$$\|A - \hat{A}\|_{2,F} = \min_{\mathrm{rank}(\tilde{A}) \le p} \|A - \tilde{A}\|_{2,F}. \tag{5}$$

Thus small singular values indicate that A is close to some rank deficient $\hat{A}$. This is the basis of the standard approach to rank estimation, [4]; the effectiveness of most alternate approaches is typically measured in terms of their ability to reveal the presence of small singular values. In generalizing (5) to recover the rank of $\hat{A}\hat{B}^{\mathrm{T}}$ from A and B we will estimate

$$d^2(A,B) = \min_{\mathrm{rank}(\hat{A}\hat{B}^{\mathrm{T}}) \le (m_a - 1)} \|A - \hat{A}\|_F^2 + \|B - \hat{B}\|_F^2. \tag{6}$$

Our estimate of (6) depends on a perturbation expansion and will not be exact; the estimated $d^2(A,B)$ can be larger than the true value by a quanitity that is $O(d^3(A,B))$.

In §2 we state and prove the main theorem. We outline an algorithm that attempts to compute a completely rank revealing product decomposition in §4. Given prescribed ranks r_a, r_b and r_p, the goal is to compute a decomposition that reveals nearly minimal perturbations to find nearby $\hat{A}$ and $\hat{B}$ with the prescribed product SVD structure. The performance of the algorithm as judged by this standard depends on the accuracy of the perturbation expansion given in §2 and on the accuracy of a deflation step that is used to generalize the decomposition to the case of higher order rank deficiency. Both pose difficulties which we will illustrate with examples. Nevertheless, in practice the overall method seems to be more robust than other methods for estimating the rank of a matrix product.

2 The Eckhart-Young Generalization

We assume that $\mathrm{rank}(AB^{\mathrm{T}}) = m_a$ and that A and B come from (1) for some rank deficient $\hat{A}\hat{B}^{\mathrm{T}}$. We will develop an estimate of $d^2(A,B)$ that can be computed using just A and B. The estimate is based on an expansion in terms of the perturbations E and F. To provide a compact notation for neglecting higher order terms, we will set

$$\epsilon = \max(\|E\|_F, \|F\|_F)$$

and freely ignore $O(\epsilon^2)$ terms in expressions for quantities that are $O(1)$ or $O(\epsilon)$. In some cases we will neglect terms of $O(\epsilon^3)$ in expansions of terms that are $O(\epsilon^2)$. The occasional need for a higher order expansion arises from a theorem from [5] in which second order terms are kept to retain accuracy in a perturbation expansion for a singular value that is very small or exactly zero. All vector norms will be 2-norms. For a matrix X for which XX^{T} is nonsingular we define the projection

$$P_X = X^{\mathrm{T}} \left(XX^{\mathrm{T}} \right)^{-1} X.$$

Our main result, of which most of this section is an extended proof, is the following theorem.

Theorem 1 *For A and B given by perturbations*

$$A = \hat{A} + E, \qquad B = \hat{B} + F$$

we assume that AB^{T} has full rank and that $\hat{A}\hat{B}^{\mathrm{T}}$ has rank $m_a - 1$. For some u satisfying $\|u\| = 1$ let

$$y = \frac{A^{\mathrm{T}}u}{\|A^{\mathrm{T}}u\|},$$

$$g = \left(P_A^{\perp}B^{\mathrm{T}}BP_A^{\perp} + \|A^{\mathrm{T}}u\|^2 I_n\right)^{-1} P_A^{\perp}B^{\mathrm{T}}BA^{\mathrm{T}}u$$

and

$$\begin{aligned} f &= B\left(y - \frac{1}{\|A^{\mathrm{T}}u\|}g\right) \\ &= \|A^{\mathrm{T}}u\|\left(BP_A^{\perp}B^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}u. \end{aligned}$$

Then

$$u^{\mathrm{T}}(A - ug^{\mathrm{T}})(B - fy^{\mathrm{T}})^{\mathrm{T}} = 0. \tag{7}$$

If u satisfies

$$\lambda_{m_a}\left(AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}\right) = u^{\mathrm{T}}AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}u \tag{8}$$

where $\lambda_{m_a}(\cdot)$ is the smallest eigenvalue (singular value) of the $m_a \times m_a$ matrix and $\|u\| = 1$ and $\hat{A}$ has full rank then for sufficiently small perturbations E and F

$$\begin{aligned} \|ug^{\mathrm{T}}\|_{2,F}^2 + \|yf^{\mathrm{T}}\|_{2,F}^2 &= \lambda_{m_a}\left(AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}\right) && (9) \\ &\leq (\|E\|_F + \|F\|_F)^2 + O(\epsilon^3). \blacksquare && (10) \\ & && (11) \end{aligned}$$

A solution to (8) with $\|u\| = 1$ always exists.

The theorem shows that $A - ug^{\mathrm{T}}$ and $B - fy^{\mathrm{T}}$ are a pair with a rank deficient product that is nearly as close as possible to A and B in the sense of (6).

The conditions that make (10) valid are hard to state in a more precise form. This and the fact that the phrase "sufficiently small" must be interpreted in terms of the unknown matrices $\hat{A}$ and $\hat{B}$ pose significant problems in evaluating the quality of the expansion. Nevertheless, experiments suggest that (10) is often a significant improvement over other methods for finding rank deficiency in a matrix product.

The theorem makes five claims: the existence of a solution to (8), the equivalence of the two relations for f, the fact that u is a null vector of the perturbed matrix pair in (7), the equivalence of the eigenvalue λ_{m_a} with the sums of the norms of the perturbations in (9) and the upper bound (10). The distance estimate (10) is the most difficult to verify. The presence of the term $\|A^{\mathrm{T}}u\|^2 I_{m_b}$ in (8) means that proving the existence of a solution u is

also a nontrivial matter. We will deal with these two issues after proving the simpler parts of the theorem.

To prove (9) we will show that the equality

$$\|ug^{\mathrm{T}}\|_{2,F}^2 + \|fy^{\mathrm{T}}\|_{2,F}^2 = u^{\mathrm{T}}AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}u$$

holds for any u with $\|u\| = 1$. Since (8) holds by assumption this implies (9). To complete this proof we note that as it is defined in the theorem g is the solution to the least squares problem

$$\min_g \left\| \begin{bmatrix} \frac{1}{\|A^{\mathrm{T}}u\|} BP_A^{\perp} \\ I_n \end{bmatrix} g - \begin{bmatrix} \frac{1}{\|A^{\mathrm{T}}u\|} BA^{\mathrm{T}}u \\ 0 \end{bmatrix} \right\|^2 .$$

From the presence of the projection $P_A^{\perp}$ in the first m_b rows of the least squares problem and the penalty on the norm of g in the last n rows, it follows that $P_A^{\perp}g = g$. Since $\|u\| = \|y\| = 1$ and since $P_A^{\perp}g = g$, the residual of the least squares problem is

$$\left\| \frac{1}{\|A^{\mathrm{T}}u\|} BP_A^{\perp}g - \frac{1}{\|A^{\mathrm{T}}u\|} BA^{\mathrm{T}}u \right\|^2 + \|g\|^2 = \left\| \frac{1}{\|A^{\mathrm{T}}u\|} Bg - By \right\|^2 + \|g\|^2 = \|ug^{\mathrm{T}}\|_{2,F}^2 + \|fy^{\mathrm{T}}\|_{2,F}^2.$$

To make the connection with the right hand side of (9), we can get an alternate formula for this residual using the orthogonality property of least squares solutions. In particular

$$\begin{aligned}
\|ug^{\mathrm{T}}\|_{2,F}^2 + \|fy^{\mathrm{T}}\|_{2,F}^2 &= \begin{bmatrix} \frac{1}{\|A^{\mathrm{T}}u\|} u^{\mathrm{T}}AB^{\mathrm{T}} & 0 \end{bmatrix} \left(\begin{bmatrix} \frac{1}{\|A^{\mathrm{T}}u\|} BA^{\mathrm{T}}u \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{\|A^{\mathrm{T}}u\|} BP_A^{\perp} \\ I_n \end{bmatrix} g \right) \\
&= u^{\mathrm{T}}AB^{\mathrm{T}} \left(\frac{1}{\|A^{\mathrm{T}}u\|^2} I_{m_b} - \frac{BP_A^{\perp}}{\|A^{\mathrm{T}}u\|^2} \left(\frac{1}{\|A^{\mathrm{T}}u\|^2} P_A^{\perp}B^{\mathrm{T}}BP_A^{\perp} + \right. \right. \\
&\quad \left. \left. I_n \right)^{-1} \frac{P_A^{\perp}B^{\mathrm{T}}}{\|A^{\mathrm{T}}u\|^2} \right) BA^{\mathrm{T}}u \\
&= u^{\mathrm{T}}AB^{\mathrm{T}} \left(BP_A^{\perp}B^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b} \right)^{-1} BA^{\mathrm{T}}u.
\end{aligned}$$

The final equality follows from the Sherman-Morrison-Woodbury matrix inversion formula

$$\left(X + YZ^{\mathrm{T}}\right)^{-1} = X^{-1} - X^{-1}Y\left(I + Z^{\mathrm{T}}X^{-1}Y\right)^{-1} Z^{\mathrm{T}}X^{-1}.$$

Together with (8) this establishes (9).

Verification that the two equations for f are equivalent is by use of $P_A^{\perp}g = g$, substitution of the expressions for y and g and another application of the Sherman-Morrison-Woodbury formula. By harmlessly inserting $P_A^{\perp}$ into the first formula for f we get

$$\begin{aligned}
f &= B\left(y - \frac{1}{\|A^{\mathrm{T}}u\|} P_A^{\perp}g\right) \\
&= \|A^{\mathrm{T}}u\| \left(\frac{1}{\|A^{\mathrm{T}}u\|^2} I_{m_b} - \frac{BP_A^{\perp}}{\|A^{\mathrm{T}}u\|^2} \left(\frac{1}{\|A^{\mathrm{T}}u\|^2} P_A^{\perp}B^{\mathrm{T}}BP_A^{\perp} + I_n \right)^{-1} \frac{P_A^{\perp}B^{\mathrm{T}}}{\|A^{\mathrm{T}}u\|^2} \right) BA^{\mathrm{T}}u \\
&= \|A^{\mathrm{T}}u\| \left(BP_A^{\perp}B^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b} \right)^{-1} BA^{\mathrm{T}}u.
\end{aligned}$$

We will now show that u is a left null vector of the perturbed matrix product. Since $P_A^{\perp} g = g$ and $y \in R(A^{\mathrm{T}})$, $y^{\mathrm{T}} g = 0$ and

$$\begin{aligned} u^{\mathrm{T}}(A - ug^{\mathrm{T}})(B - fy^{\mathrm{T}})^{\mathrm{T}} &= u^{\mathrm{T}}AB^{\mathrm{T}} - g^{\mathrm{T}}B^{\mathrm{T}} - u^{\mathrm{T}}Ay\left(y^{\mathrm{T}} - \frac{1}{\|A^{\mathrm{T}}u\|}g^{\mathrm{T}}\right)B^{\mathrm{T}} \\ &= u^{\mathrm{T}}A(I - yy^{\mathrm{T}})B^{\mathrm{T}} - g^{\mathrm{T}}B^{\mathrm{T}} + \frac{u^{\mathrm{T}}Ay}{\|A^{\mathrm{T}}u\|}g^{\mathrm{T}}B^{\mathrm{T}} = 0. \end{aligned}$$

The cancellations happen because $u^{\mathrm{T}}Ay = \|A^{\mathrm{T}}u\|$ and

$$u^{\mathrm{T}}A(I - yy^{\mathrm{T}}) = \|A^{\mathrm{T}}u\| y^{\mathrm{T}} P_{y^{\mathrm{T}}}^{\perp} = 0.$$

To show that a solution to (8) exists define

$$\lambda_{m_a}(\alpha) = \lambda_{m_a}\left(AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \alpha^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}\right)$$

for $\alpha > 0$. Let $u_{m_a}(\alpha)$ be an eigenvector associated with $\sigma_{m_a}(\alpha)$. We need to show that there is an α_0 such that

$$\|A^{\mathrm{T}} u_{m_a}(\alpha_0)\| = \alpha_0 \tag{12}$$

for some choice of the singular vector $u_{m_a}(\alpha_0)$.

If $u_{m_a}(\alpha)$ were unique (up to sign change) for each α and continuous as a function of α, then the proof would be very simple. Since A by assumption has full rank

$$f(\alpha) = \|A^{\mathrm{T}} u_{m_a}(\alpha)\| - \alpha > 0$$

for sufficiently small α. For sufficiently large α, $f(\alpha) < 0$. Hence continuity would guarantee a solution for which $f(\alpha) = 0$. Unfortunately a rigorous proof is complicated by possible discontinuities in $u_{m_a}(\alpha)$ that can occur when the eigenvalue $\lambda_{m_a}(\alpha)$ is repeated. The proof of the following result uses only basic ideas from analysis together with the well known fact that a particular eigenvalue of a family of matrices varying continuously with $\alpha > 0$ is also continuous in α.

Theorem 2 *If A has full rank then (8) is satisfied for some u normalized so that $\|u\| = 1$.*

Proof: As we have noted, we need to show that (12) holds for some $\alpha_0 > 0$. Since A has full rank

$$f(\alpha) = \min_{u_{m_a}} \|A^{\mathrm{T}} u_{m_a}(\alpha)\| - \alpha > 0$$

for all sufficiently small $\alpha > 0$. We also have $f(\alpha) < 0$ for all sufficiently large $\alpha > 0$. The minimum that defines $f(\alpha)$ is taken over all possible choices of $u_{m_a}(\alpha)$ (i.e. all norm one vectors in the subspace spanned by the eigenvectors associated with $\lambda_{m_a}(\alpha)$). Let $\alpha_0 > 0$ be defined by

$$\alpha_0 = \sup\{\alpha \mid f(\alpha) \geq 0\}.$$

From this definition it follows that either $f(\alpha_0) \geq 0$ or in any interval $[\alpha_0 - \eta, \alpha_0]$ there must be an infinite number of points for which $f(\alpha_0) \geq 0$. Either way it follows that there

exist sequences $\overline{u}_k$ and $\overline{\alpha}_k$ such that$\|A^{\mathrm{T}}\overline{u}_k\| \geq \overline{\alpha}_k$ where $\overline{u}_k$ is an eigenvector associated with $\lambda_{m_a}(\overline{\alpha}_k)$, $\overline{\alpha}_k \leq \alpha_0$ for all k and $\overline{\alpha}_k \to \alpha_0$. The continuity of eigenvalues implies that the bounded sequence $\overline{u}_k$ has some subsequence converging to $\overline{u}$ such that $\|A^{\mathrm{T}}\overline{u}\| \geq \alpha_0$ where $\overline{u}$ is an eigenvector associated with $\lambda_{m_a}(\alpha_0)$.

Similarly the definition of α_0 implies that $f(\alpha) < 0$ for $\alpha > \alpha_0$. From this the same argument used to construct $\overline{u}$ implies that there is an eigenvector $\underline{u}$ such that

$$\|A^{\mathrm{T}}\underline{u}\| \leq \alpha_0 \leq \|A^{\mathrm{T}}\overline{u}\|.$$

It follows that there are scalars c and s such that $c^2 + s^2 = 1$ and $u = c\underline{u} + s\overline{u}$ is also an eigenvector associated with $\lambda_{m_a}(\alpha_0)$ and $\|A^{\mathrm{T}}u\| = \alpha_0$. ■

At this point we have verified all of Theorem 1 except for (10). To finish the proof, we start with the following perturbation expansion for singular values from [5].

Theorem 3 *For a general $m_a \times m_b$ matrix $\hat{D}$ with SVD*

$$U^{\mathrm{T}}\hat{D}V = \begin{bmatrix} U_1^{\mathrm{T}} \\ u_2^{\mathrm{T}} \end{bmatrix} \hat{D} \begin{bmatrix} V_1 & v_2 & V_3 \end{bmatrix} = \begin{bmatrix} \hat{\Sigma} & 0 & 0 \\ 0 & \hat{\sigma}_{m_a} & 0 \end{bmatrix}$$

with $U^{\mathrm{T}}U = I_{m_a}$, $V^{\mathrm{T}}V = I_{m_b}$ and invertible $\hat{\Sigma} - \hat{\sigma}_{m_a}I_{m_a-1}$ let

$$U^{\mathrm{T}}HV = \begin{bmatrix} G_{11} & g_{12} & G_{13} \\ g_{21}^{\mathrm{T}} & \gamma_{22} & g_{23}^{\mathrm{T}} \end{bmatrix}$$

and $h = \hat{\sigma}_{m_a}g_{12} + \Sigma g_{21}$. Then $D = \hat{D} + H$ has a singular value

$$\sigma_{m_a}^2 = (\hat{\sigma}_{m_a} + \gamma_{22})^2 + \|g_{21}\|^2 + \|g_{23}\|^2 + h^{\mathrm{T}}(\hat{\sigma}_{m_a}^2 I_{m_a-1} - \Sigma^2)^{-1}h + O(\|E\|^3).$$ ■

By considering second order terms this expansion can accurately characterize the effect of a perturbation on a zero singular value $\hat{\sigma}_{m_a} = 0$ so long as $\hat{\Sigma}$ has full rank. In particular, if $\hat{\sigma}_{m_a} = 0$ then

$$\sigma_{m_a}^2 = \gamma_{22}^2 + \|g_{23}\|^2. \tag{13}$$

By assumption $\hat{A}\hat{B}^{\mathrm{T}}$ has rank $m_a - 1$. If we define

$$\hat{C} = \hat{A}\hat{B}^{\mathrm{T}}\left(\hat{B}P_{\hat{A}}^{\perp}\hat{B}^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b}\right)^{-1}\hat{B}\hat{A}^{\mathrm{T}}$$

for some particular choice of u satisfying (8) and also define

$$\hat{D} = \hat{A}\hat{B}^{\mathrm{T}}\left(\hat{B}P_{\hat{A}}^{\perp}\hat{B}^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b}\right)^{-1/2}$$

then both $\hat{C}$ and $\hat{D}$ have rank $m_a - 1$. Consequently $\hat{\Sigma}$ as defined in terms of $\hat{D}$ in Theorem 3 has full rank. This shows that the theorem can be applied to estimate the effect of the perturbations E and F on $\hat{\sigma}_{m_a} = \sigma_{m_a}(\hat{D})$.

More precisely, if the square root is defined as any matrix $\hat{D}$ such that

$$\hat{D}\hat{D}^{\mathrm{T}} = \hat{C}$$

and if we also define

$$D = AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b}\right)^{-1/2}$$

then we will derive an expansion of the form

$$D = \hat{D} + H + O(\epsilon^2) \tag{14}$$

where H is a perturbation defined in terms of E and F. Note that in defining D and $\hat{D}$ we have used the quantity $\|A^{\mathrm{T}}u\|$ with the perturbed A in both cases. Using the expression we will derive for H and Theorem 3 we will show that if u satisfies (8) then

$$\lambda_{m_a}(C) = \sigma_{m_a}^2(D) \leq (\|E\|_F + \|F\|_F)^2 + O(\epsilon^3)$$

where C is defined in terms of A and B in a manner analogous to the definition of $\hat{C}$. This will complete the proof of (10).

The matrix square root as we have defined it is nonunique. However since $\sigma_{m_a}(D)$ does not depend on the choice of square root, we are free to choose particular square roots D and $\hat{D}$ to make H suitably small subject only to the constraints $DD^{\mathrm{T}} = C$ and $\hat{D}\hat{D}^{\mathrm{T}} = \hat{C}$. We start by choosing an arbitrary factorization of the form

$$\hat{S}\hat{S}^{\mathrm{T}} = \left(\hat{B}P_{\hat{A}}^{\perp}\hat{B}^{\mathrm{T}} + \|A^{\mathrm{T}}u\|I_{m_b}\right)^{-1}.$$

This is guaranteed to exist since the inverse matrix is symmetric and positive definite. We let

$$\hat{D} = \hat{A}\hat{B}^{\mathrm{T}}\hat{S}.$$

For sufficiently small E and F perturbing $\hat{A}$ and $\hat{B}$ we can choose a square root S such that

$$S = \hat{S} + H_S + O(\epsilon^2)$$

where $\|H_S\| = O(\epsilon)$. It turns out that the precise magnitude of $\|H_S\|$ is of no consequence to the analysis; all that matters is that terms of the form $O(\epsilon H_S)$ are $O(\epsilon^2)$ and are consequently negligible.

We can expand D in terms of E, F and H_S as follows

$$\begin{aligned}
D &= \left(\hat{A}+E\right)\left(\hat{B}+F\right)^{\mathrm{T}}\left(\hat{S}+H_S\right) \\
&= \hat{A}\hat{B}^{\mathrm{T}}\hat{S} + E\hat{B}^{\mathrm{T}}\hat{S} + \hat{A}F^{\mathrm{T}}\hat{S} + \hat{A}\hat{B}^{\mathrm{T}}H_S + O(\epsilon^2) \\
&= \hat{A}\hat{B}^{\mathrm{T}}\hat{S} + EP_{\hat{A}}\hat{B}^{\mathrm{T}}\hat{S} + EP_{\hat{A}}^{\perp}\hat{B}^{\mathrm{T}}\hat{S} + \hat{A}F^{\mathrm{T}}\hat{S} + \hat{A}\hat{B}^{\mathrm{T}}H_B + O(\epsilon^2) \\
&= \hat{A}\hat{B}^{\mathrm{T}}\hat{S} + H.
\end{aligned}$$

By the assumption that $\mathrm{rank}(\hat{A}\hat{B}^{\mathrm{T}}) = m_a$, the matrix $\hat{D} = \hat{A}\hat{B}^{\mathrm{T}}\hat{S}$ has rank $m_a - 1$ and a unique left null vector u_2 that depends only on the left null space of $\hat{A}\hat{B}^{\mathrm{T}}$ and not on $\hat{S}$ or the value of $\|A^{\mathrm{T}}u\|$.

We consider Theorem 3 applied to the smallest singular value $\sigma_{m_a}(\hat{D}) = 0$ with the perturbation H. Using (13) we get

$$\sigma_{m_a}^2(D) = (u_2^{\mathrm{T}}Hv_2)^2 + \|u_2^{\mathrm{T}}HV_3\|^2 + O(\epsilon^3)$$

where u_2 is the left null vector of $\hat{D}$.

Note that

$$P_{\hat{A}}\hat{B}^{\mathrm{T}}\hat{S}\begin{bmatrix} v_2 & V_3 \end{bmatrix} = \hat{A}^{\mathrm{T}}\left(\hat{A}\hat{A}^{\mathrm{T}}\right)^{\mathrm{T}}\hat{D}\begin{bmatrix} v_2 & V_3 \end{bmatrix} = 0$$

so that many of the terms in the expansion for H share either common left or common right null vectors with $\hat{D}$. Consequently

$$\begin{aligned} \sigma_{m_a}^2(D) &= \left\| u_2^{\mathrm{T}}(\hat{A}F^{\mathrm{T}}\hat{S} + EP_{\hat{A}}^{\perp}\hat{B}^{\mathrm{T}}\hat{S})\begin{bmatrix} v_2 & V_3 \end{bmatrix}\right\|^2 + O(\epsilon^3) \\ &\leq \left(\|\hat{A}^{\mathrm{T}}u_2\|\|\hat{S}\|_2\|F\|_2 + \|P_{\hat{A}}^{\perp}\hat{B}^{\mathrm{T}}\hat{S}\|_2\|E\|_2\right)^2 + O(\epsilon^3). \end{aligned}$$

For E and F that are sufficiently small the assumption that $\mathrm{rank}(\hat{A}\hat{B}^{\mathrm{T}}\hat{S}) = \mathrm{rank}(\hat{A}\hat{B}) = m_a - 1$ guarantees that u is a perturbed version of the left null vector u_2. Since $u = u_2 + O(\epsilon)$ we have

$$\|A^{\mathrm{T}}u_2\| = \|A^{\mathrm{T}}u\| + O(\epsilon).$$

The qualification that E and F must be sufficiently small is standard in any first order perturbation expansion of a singular vector associated with an isolated singular value.

The fact that $\|\hat{S}\|_2 \leq 1/\|A^{\mathrm{T}}u\| + O(\epsilon)$ guarantees that

$$\|A^{\mathrm{T}}u_2\|\|\hat{S}\|_2 \leq 1 + O(\epsilon).$$

Similarly

$$\|P_{\hat{A}}^{\perp}\hat{B}^{\mathrm{T}}\hat{S}\|_2 \leq 1 + O(\epsilon).$$

Thus

$$\begin{aligned} \sigma_{m_a}^2(D) &\leq \left(\|E\|_F + \|F\|_F\right)^2 + O(\epsilon^3) \\ &\leq 2\left(\|E\|_F^2 + \|F\|_F^2\right) + O(\epsilon^3). \end{aligned} \tag{15}$$

Since $\sigma_{m_a}^2(D) = \lambda_{m_a}(C)$, at this point we have proven all the claims of Theorem 1.

We can offer further justification for the assumption that $\hat{A}\hat{B}^{\mathrm{T}}$ has rank $m_a - 1$. Instead of letting E and F be arbitrary perturbations, we define

$$(E,F) = \underset{\{(E,F)\,\big|\,\mathrm{rank}((A+E)(B+F)^{\mathrm{T}}) < m_a\}}{\mathrm{argmin}} \|E\|_F^2 + \|F\|_F^2. \tag{16}$$

Thus

$$d^2(A,B) = \|E\|_F^2 + \|F\|_F^2.$$

With $\hat{A}$ and $\hat{B}$ defined by (1), the matrix pair $\hat{A}$ and $\hat{B}$ is a closest pair to A and B for which $\hat{A}\hat{B}^{\mathrm{T}}$ is rank deficient.

Theorem 4 *For AB^{T} with rank m_a and E and F given by (16), with $\hat{A}$ and $\hat{B}$ defined by (1), the rank of $\hat{A}\hat{B}^{\mathrm{T}}$ is exactly $m_a - 1$.*

Proof: Any degree of rank deficiency in $\hat{A}\hat{B}^{\mathrm{T}}$ implies that there exists a vector u such that $u^{\mathrm{T}}\hat{A}\hat{B}^{\mathrm{T}} = 0$ and $\|u\| = 1$. This in turn implies that orthogonal transformations U and W can be constructed to give

$$U^{\mathrm{T}}\hat{A}Q = \begin{bmatrix} \hat{a}_{11} & 0 \\ \hat{a}_{21} & \hat{A}_{22} \end{bmatrix}, \qquad \hat{B}Q = \begin{bmatrix} 0 & \hat{B}_2 \end{bmatrix} \tag{17}$$

where $Ue_1 = u$ and a_{11} is a scalar. Since perturbing the nonzero elements of (17) only increases the Frobenius norm of the perturbations it follows from the assumed minimality of the perturbations that

$$U^{\mathrm{T}}AQ = \begin{bmatrix} \hat{a}_{11} & e^{\mathrm{T}} \\ \hat{a}_{21} & \hat{A}_{22} \end{bmatrix}, \qquad BQ = \begin{bmatrix} f & \hat{B}_2 \end{bmatrix}$$

and

$$E = U \begin{bmatrix} 0 & e^{\mathrm{T}} \\ 0 & 0 \end{bmatrix} Q^{\mathrm{T}}, \qquad F = \begin{bmatrix} f & 0 \end{bmatrix} Q^{\mathrm{T}}. \tag{18}$$

These perturbations induce a higher order rank deficiency in $\hat{A}\hat{B}^{\mathrm{T}}$ only if

$$\mathrm{rank}(\hat{A}_{22}\hat{B}_2^{\mathrm{T}}) < m_a - 1.$$

However if that is the case and $e \neq 0$, then a strictly smaller pair of perturbations

$$E = 0, \qquad F = \begin{bmatrix} f & 0 \end{bmatrix} Q^{\mathrm{T}}$$

gives $\mathrm{rank}(\hat{A}\hat{B}^{\mathrm{T}}) < m_a$. This contradicts the assumed minimality of E and F. On the other hand, if $e = 0$ then

$$\hat{A}\hat{B}^{\mathrm{T}} = AB^{\mathrm{T}} - AQ \begin{bmatrix} f & 0 \end{bmatrix}^{\mathrm{T}}$$

so that the perturbation to AB^{T} is rank one and cannot cause the rank of $\hat{A}\hat{B}^{\mathrm{T}}$ to be less than $m_a - 1$. ■

It follows from this theorem that the closest pair to A and B with a rank deficient product has $\mathrm{rank}(\hat{A}\hat{B}^{\mathrm{T}}) = m_a - 1$. This pair can serve as the $\hat{A}$ and $\hat{B}$ mentioned in Theorem 1 giving

$$\|ug^{\mathrm{T}}\|_{2,F}^2 + \|yf^{\mathrm{T}}\|_{2,F}^2 \leq d^2(A,B) + O(\epsilon^3).$$

While the observation that $\hat{A}\hat{B}^{\mathrm{T}}$ defined in this way always has rank $m_a - 1$ is comforting, it doesn't really give obvious benefit in evaluating the quality of the expansion (10).

The following example shows, the expansion fail dramatically if $\hat{A}$ does not have full rank.

Example 2 Consider the matrix pair

$$A = \begin{bmatrix} \epsilon & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{19}$$

for $0 < \epsilon \ll 1$. While we do not rigorously prove the fact, it is easy to show that the matrix pair that is closest to A and B in the sense of (6) is

$$\hat{A} = \begin{bmatrix} 0 & 0 \end{bmatrix}, \qquad \hat{B} = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Even without proof this is quite plausible: any perturbation to the zero elements of A or B results in a negligible $O(\epsilon^2)$ change to $\hat{A}\hat{B}^{\mathrm{T}}$. That Theorem 1 fails follows from the fact that for $m_a = 1$, we have $u = 1$ and

$$\sigma_{m_a}\left(AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \|A^{\mathrm{T}}u\|^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}\right) = \epsilon(0 + \epsilon^2 I_{m_b})^{-1}\epsilon = 1. \blacksquare$$

The method of looking at the smallest singular value of $B_1^{(1)}$, $\sigma_1(B_1^{(1)}) = 1$, fails in exactly the same way on (19). However in less contrived problems, the expansion in Theorem 1 often retains its accuracy even when A is very ill-conditioned and the smallest singular value of $B_1^{(1)}$ is orders of magnitude larger than $d(A, B)$. It is only in the higher order terms that ill-conditioning in A can have an effect on (10). In contrast, the effect of errors on the smallest singular value of $B_1^{(1)}$ is often directly magnified by a factor proportional to the condition number of A.

3 Finding the Null Vector

Theorem 1 characterizes nearly minimal perturbations in terms of a possibly nonunique solution to the nonlinear equation (8). In contrast to a direct optimization formulation in which we might have to worry about local minima, the perturbation analysis of the last section shows that any solution to (8) will give suitably small perturbations so long as the conditions on the validity of the expansion are not too close to being violated. Subject to these conditions, purely local information obtained by solving (8) provides an estimate of a globally minimum perturbations giving rank deficiency.

However solving (8) is not always a trivial task. The singular vector associated with the singular value

$$\lambda_{m_a}\left(AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \alpha^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}\right)$$

is not in general continuous as a function of the parameter α at points where the singular value has multiplicity greater than one. Because of this potential discontinuity, a provably convergent iteration for solving (8) seems to be a remote possibility. Nevertheless, the left singular vector associated with $\sigma_{m_a}(AB^{\mathrm{T}})$ is often a good approximation to u. Making use of this starting vector, we propose the following iteration without any guarantees of convergence.

Algorithm 1 Take u_0 to be the left singular vector of AB^{T} associated with $\sigma_{m_a}(AB^{\mathrm{T}})$. Let $\alpha_0 = \|A^{\mathrm{T}}u_0\|$ and let $k = 1$.

1. Let u_k be the eigenvector vector defined by

$$u_k^{\mathrm{T}}AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \alpha_{k-1}^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}u_k = \lambda_{m_a}\left(AB^{\mathrm{T}}\left(BP_A^{\perp}B^{\mathrm{T}} + \alpha_{k-1}^2 I_{m_b}\right)^{-1} BA^{\mathrm{T}}\right).$$

2. Let $\alpha_k = \|A^{\mathrm{T}}u_k\|$.

3. $k \leftarrow k + 1$. Go to 1.

In most cases, the algorithm converges to a solution of (8). However it does not always work. The following example illustrates the potential problems that can arise with repeated singular values.

Example 3 Let

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 3 \end{bmatrix}.$$

It is easy to verify that the application of Algorithm 1 to this problem gives

$$(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \ldots) = (1, 2, 1, 2, \ldots).$$

The algorithm does not converge, but a solution to (8) occurs for $\alpha = \sqrt{5/3}$ for which

$$AB^{\mathrm{T}} \left(BP_A^{\perp} B^{\mathrm{T}} + \alpha^2 I_{m_b} \right)^{-1} BA^{\mathrm{T}} = \begin{bmatrix} 3/8 & 0 \\ 0 & 3/8 \end{bmatrix}$$

has a repeated singular value and we can choose

$$u = \begin{bmatrix} \sqrt{7}/3 & \sqrt{2}/3 \end{bmatrix}$$

to satisfy $\|A^{\mathrm{T}} u\|^2 = 5/3 = \alpha^2$. ■

The example is somewhat contrived. Algorithm 1 seems to converge in most cases.

4 Higher Order Rank Deficiency

If A and B come from a model of the form (1) with $\mathrm{rank}(\hat{A}\hat{B}^{\mathrm{T}}) = r < m_a - 1$ we might wish to find perturbations E and F that satisfy

$$(E, F) = \underset{\{(E,F) \,|\, \mathrm{rank}((A+E)(B+F)^{\mathrm{T}}) \leq r\}}{\mathrm{argmin}} \|E\|_F^2 + \|F\|_F^2. \tag{20}$$

The most obvious methods for attempting to decide if A and B are consistent with a model of the form (1) with a rank r matrix product are natural generalizations of methods for the case $r = m_a - 1$. As discussed in §1 we can attempt to determine the number of non-negligible singular values of either AB^{T} or $B_1^{(1)}$ where $B_1^{(1)}$ is defined as in (3). The potential problems with these rank decisions are the same as before.

Theorem 1 does not admit an obvious generalization to $r < m_a - 1$. As an alternative we will assume that perturbations that further reduce the rank $m_a - 1$ product to rank r can be chosen to be orthogonal to the perturbations that originally reduced the product to rank $m_a - 1$. This assumption is inspired by the SVD and the orthogonality properties of the minimal perturbations giving rank deficiency in a single matrix. While it is not correct in the case of a matrix product, some such assumption seems necessary to permit recursive application of Theorem 1. For this reason neither the theoretical nor practical results of this section will be completely satisfactory. We will not be able to guarantee that the method

reveals the rank of AB^{T} in the sense of finding nearly minimal E and F. Nevertheless, we will get an algorithm that typically finds smaller perturbations than can be found by looking for small singular values in $B_1^{(1)}$.

The distance measure (6) is invariant under orthogonal transformations. Thus we are free to apply transformations of the form $U^{\mathrm{T}}AQ$ and $V^{\mathrm{T}}BQ$ where U, V and Q are orthogonal. We start by with perturbations $\tilde{E}^{(1)}$ and $\tilde{F}^{(1)}$ and a vector u such that

$$u^{\mathrm{T}}\left(A+\tilde{E}^{(1)}\right)\left(B+\tilde{F}^{(1)}\right)^{\mathrm{T}}=0$$

with $\|u\|=1$ and $\|\tilde{E}^{(1)}\|_F^2+\|\tilde{F}^{(1)}\|_F^2$ not much larger than $d^2(A,B)$. Our goal is to find further perturbations $E^{(2)}$ and $F^{(2)}$ so that

$$\operatorname{rank}\left(\left(A+\tilde{E}^{(1)}+E^{(2)}\right)\left(B+\tilde{F}^{(1)}+F^{(2)}\right)^{\mathrm{T}}\right)=r \tag{21}$$

with nearly minimal $\left\|\tilde{E}^{(1)}+E^{(2)}\right\|_F^2+\left\|\tilde{F}^{(1)}+F^{(2)}\right\|_F^2$.

Before describing our orthogonality assumption and attempting to construct $E^{(2)}$ and $F^{(2)}$ we consider the computation of $\tilde{E}^{(1)}$ and $\tilde{F}^{(1)}$. Theorem 1 gives an explicit formula for nearly minimal perturbations. However we can do slightly better. Instead of using Theorem 1 directly, we will use u solving (8) to construct alternate perturbations that can be slightly smaller than those of the theorem. These refined rank reducing perturbations also have the advantage of more directly illuminating the algorithmic significance of the assumption of orthogonality between the perturbations.

Suppose that u, g, y and f are defined as in the theorem. If

$$E^{(1)}=-ug^{\mathrm{T}},\qquad F^{(1)}=-fy^{\mathrm{T}} \tag{22}$$

are the rank reducing perturbations then the theorem states that

$$u^{\mathrm{T}}\left(A+E^{(1)}\right)\left(B+F^{(1)}\right)^{\mathrm{T}}=0.$$

Choose an orthogonal Q such that

$$u^{\mathrm{T}}\left(A+E^{(1)}\right)Q=u^{\mathrm{T}}\left(A+E^{(1)}\right)\begin{bmatrix}q_1 & Q_2\end{bmatrix}=\begin{bmatrix}\beta & 0^{\mathrm{T}}\end{bmatrix}$$

for some scalar β and an orthogonal U so that $U^{\mathrm{T}}u=e_1$ where e_1 is the first standard basis vector. Then

$$U^{\mathrm{T}}AQ=\begin{bmatrix}\tilde{a}_{11} & \tilde{g}\\ \tilde{a}_{21} & \tilde{A}_{22}\end{bmatrix}$$

and

$$BQ=\begin{bmatrix}\tilde{f} & \tilde{B}_2\end{bmatrix}$$

where $\tilde{a}_{11}$ is a scalar, $\tilde{f}$ is a column vector. By the construction of Q

$$\|\tilde{g}\|_2=\left\|u^{\mathrm{T}}E^{(1)}Q_2\right\|_2\le\|ug^{\mathrm{T}}\|_F.$$

Also we have

$$q_1^{\mathrm{T}} \left(B + F^{(1)}\right)^{\mathrm{T}} = 0$$

so that

$$\|\tilde{f}\|_2 = \left\|F^{(1)} q_1\right\|_2 \leq \|f y^{\mathrm{T}}\|_F.$$

Thus if

$$\tilde{E}^{(1)} = U \begin{bmatrix} 0 & -\tilde{g}^{\mathrm{T}} \\ 0 & 0 \end{bmatrix} Q^{\mathrm{T}}, \qquad \tilde{F}^{(1)} = \begin{bmatrix} -\tilde{f} & 0 \end{bmatrix} Q^{\mathrm{T}} \tag{23}$$

then

$$u^{\mathrm{T}} \left(A + \tilde{E}^{(1)}\right) \left(B + \tilde{F}^{(1)}\right)^{\mathrm{T}} = 0$$

and

$$\left\|\tilde{E}^{(1)}\right\|_F^2 + \left\|\tilde{F}^{(1)}\right\|_F^2 \leq \left\|E^{(1)}\right\|_F^2 + \left\|F^{(1)}\right\|_F^2.$$

We will use the perturbations $\tilde{E}^{(1)}$ and $\tilde{F}^{(1)}$ to induce rank deficiency instead of $E^{(1)}$ and $F^{(1)}$.

The following lemma shows that if the conditions that make the expansion from Theorem 1 valid are satisfied then both these perturbations are non-zero.

Lemma 1 *Let u, g, f and y be defined as in Theorem 1, $E^{(1)}$ and $F^{(1)}$ defined by (22) and $\tilde{E}^{(1)}$ and $\tilde{F}^{(1)}$ defined by (23). The assumption*

$$u^{\mathrm{T}} A B^{\mathrm{T}} \neq 0 \tag{24}$$

implies that $\tilde{E}^{(1)} \neq 0$ and $\tilde{F}^{(1)} \neq 0$.

Proof: We extend our assumption by noting that it immediately implies $u^{\mathrm{T}} A \neq 0$ and also

$$u^{\mathrm{T}} \left(A + E^{(1)}\right) \neq 0.$$

The latter follows because Theorem 1 implies $P_A^{\perp} g = g$ so that

$$R\left(A^{\mathrm{T}}\right) \perp R\left(E^{(1)^{\mathrm{T}}}\right)$$

and $u^{\mathrm{T}} \left(A + E^{(1)}\right) = 0$ only if $u^{\mathrm{T}} A = 0$.

To prove the lemma, it is sufficient to prove that $\tilde{f} \neq 0$ and $\tilde{g} \neq 0$. We have $\tilde{f} = B q_1$ and

$$\begin{aligned} B q_1 &= \pm \frac{1}{\left\|\left(A + E^{(1)}\right)^{\mathrm{T}} u\right\|} \left(B A^{\mathrm{T}} u - B g\right) \\ &= \pm \frac{\|A^{\mathrm{T}} u\|}{\left\|\left(A + E^{(1)}\right)^{\mathrm{T}} u\right\|} f. \end{aligned}$$

By assumption $Au \neq 0$. The fact that $f \neq 0$ follows from the second expression for f in Theorem 1 and from the assumption $u^{\mathrm{T}} A B^{\mathrm{T}} \neq 0$. Thus $\tilde{f} \neq 0$.

The orthogonality of Q implies that $\tilde{g} = Q_2^{\mathrm{T}} E^{(1)^{\mathrm{T}}} u = 0$ only if $E^{(1)^{\mathrm{T}}} u = \gamma q_1$. This is equivalent to

$$E^{(1)^{\mathrm{T}}} u = \pm \frac{\gamma}{\left\| (A + E^{(1)})^{\mathrm{T}} u \right\|} \left(A + E^{(1)} \right)^{\mathrm{T}} u$$

or

$$\left(1 \mp \frac{\gamma}{\left\| (A + E^{(1)})^{\mathrm{T}} u \right\|} \right) E^{(1)^{\mathrm{T}}} u = \pm \frac{\gamma}{\left\| (A + E^{(1)})^{\mathrm{T}} u \right\|} A^{\mathrm{T}} u.$$

Since $E^{(1)^{\mathrm{T}}} u = g$, $g \perp R\left(A^{\mathrm{T}}\right)$ and $u^{\mathrm{T}} A \neq 0$ this is impossible unless $g = 0$. The fact that $g \neq 0$ follows from the definition of g and $u^{\mathrm{T}} A B^{\mathrm{T}} \neq 0$. So we must have $\tilde{g} \neq 0$. ■

Given the definition of the perturbations $\tilde{E}^{(1)}$ and $\tilde{F}^{(1)}$ we make the following assumption to allow recursive application of Theorem 1.

Assumption 1 *Given u determined as in Theorem 1 satisfying $u^{\mathrm{T}} A B^{\mathrm{T}} \neq 0$ we assume that perturbations $E^{(2)}$ and $F^{(2)}$ can be chosen so that in addition to (21) we have*

$$E^{(2)^{\mathrm{T}}} E^{(1)} = 0, \qquad F^{(2)} F^{(1)^{\mathrm{T}}} = 0 \tag{25}$$

with

$$\begin{aligned} \left\| \tilde{E}^{(1)} + E^{(2)} \right\|_F^2 + \left\| \tilde{F}^{(1)} + F^{(2)} \right\|_F^2 &= \left\| \tilde{E}^{(1)} \right\|_F^2 + \left\| E^{(2)} \right\|_F^2 + \left\| \tilde{F}^{(1)} \right\|_F^2 + \left\| F^{(2)} \right\|_F^2 \\ &\approx \min_{\mathrm{rank}(\hat{A}\hat{B}^{\mathrm{T}}) \leq r} \|A - \hat{A}\|_F^2 + \|B - \hat{B}\|_F^2. \blacksquare \end{aligned} \tag{26}$$

Lemma 1 means that $u^{\mathrm{T}} A B^{\mathrm{T}} \neq 0$ implies that the constraints (25) are nontrivial and are equivalent to

$$E^{(2)^{\mathrm{T}}} u = 0, \qquad F^{(2)} q_1 = 0. \tag{27}$$

To see how this assumption corresponds to the reduction of the order of the problem we note that with U and Q constructed as before the orthogonality relations imply that

$$E^{(2)} = U \begin{bmatrix} 0 & 0 \\ e_{21} & E_{22} \end{bmatrix} Q^{\mathrm{T}}, \qquad F^{(2)} = U \begin{bmatrix} 0 & F_2 \end{bmatrix} Q^{\mathrm{T}}.$$

Thus we have the problem of finding minimal perturbations such that

$$\begin{aligned} r &= \mathrm{rank}\left(\left(A + \tilde{E}^{(1)} + E^{(2)} \right) \left(B + \tilde{F}^{(1)} + F^{(2)} \right) \right) \\ &= \mathrm{rank}\left(\begin{bmatrix} \tilde{a}_{11} & 0 \\ \tilde{a}_{21} + e_{21} & \tilde{A}_{22} + E_{22} \end{bmatrix} \begin{bmatrix} 0 \\ \tilde{B}_2^{\mathrm{T}} + F_2^{\mathrm{T}} \end{bmatrix} \right) \\ &= \mathrm{rank}\left((\tilde{A}_{22} + E_{22})(\tilde{B}_2 + F_2)^{\mathrm{T}} \right). \end{aligned}$$

The orthogonality condition on the perturbations to B implies that e_{21} has no effect on the rank. If it is nonzero it can only increase the norm of the perturbation. Consequently we can assume that

$$E^{(2)} = U \begin{bmatrix} 0 & 0 \\ 0 & E_{22} \end{bmatrix} Q^{\mathrm{T}}$$

and try to find E_{22} and F_2 satisfying

$$\min_{\text{rank}((\tilde{A}_{22}+E_{22})(\tilde{B}_2+F_2)^{\mathrm{T}})=r} \|E_{22}\|_F^2 + \|F_2\|_F^2.$$

This is the recursive step: given the perturbations $\tilde{E}^{(1)}$ and $\tilde{F}^{(1)}$ that introduce a first order rank deficiency, we transform A and B by U and Q and continue recursively by reducing the rank of $\tilde{A}_{22}\tilde{B}_2^{\mathrm{T}}$.

Unfortunately, as we will shortly illustrate with a small example, the orthogonality and minimality conditions, (25) and (26), are not always consistent. Since it seems to be algorithmically necessary for the recursive application of Theorem 1, we will strictly enforce (25) and hope for the best in simultaneously trying to satisfy (26). The inconsistency of the conditions will degrade the ability of the algorithm to find the appropriate degree of rank deficiency in AB^{T}.

The orthogonality relations are analogous to those that apply to rank reducing perturbations given by the Eckhart-Young theorem and the SVD: if $E^{(1)}$ is a minimal perturbation such that $A + E^{(1)}$ is rank deficient, the construction of rank reducing perturbations from the SVD shows that it is possible to choose $E^{(2)}$ with ${E^{(2)}}^{\mathrm{T}}E^{(1)} = 0$ and $E^{(1)}{E^{(2)}}^{\mathrm{T}} = 0$ such that $E^{(1)} + E^{(2)}$ is a minimal perturbation reducing the rank of A to r. Our assumption (25) represent a naive attempt to extend this orthogonality property to perturbations of a matrix product.

The following example highlights the problem with this approach and with the underlying assumption that (25) is consistent with (26). For simplicity we consider an example for which rank deficiency is already present and $u^{\mathrm{T}}AB^{\mathrm{T}} = 0$. Since this implies that $\tilde{E}^{(1)} = 0$ and $\tilde{F}^{(1)} = 0$, we use the alternate form of the orthogonality constraints given in (27).

Example 4 Consider the matrices

$$A = \left[\begin{array}{c|c} A_1 & 0 \end{array}\right] = \left[\begin{array}{ccc|ccc} \delta & 0 & 0 & 0 & 0 & 0 \\ 1 & \delta & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{array}\right]$$

$$B = \left[\begin{array}{c|c} B_1 & B_2 \end{array}\right] = \left[\begin{array}{ccc|ccc} 0 & 1 & 1 & \delta & 0 & 0 \\ 0 & -\epsilon/\delta & -\epsilon & 0 & \delta & 0 \\ 0 & 0 & 0 & 0 & 0 & \delta \end{array}\right]$$

where $0 < \epsilon \ll \delta \ll 1$.

The product is already rank deficient and it is easy to verify that the vector

$$u^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

is a solution to (8). Since no perturbation is required to give rank deficiency we can choose

$$q_1^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus we can set $U = I$ and $Q = I$.

The matrix pair is close to a pair for which the product has rank 1. If

$$E_2 = \begin{bmatrix} 0 & -\epsilon & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad F_1 = \begin{bmatrix} 0 & 0 & 0 \\ \epsilon & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

then

$$\begin{bmatrix} A_1 & E_2 \end{bmatrix} \begin{bmatrix} B_1^{\mathrm{T}} + F_1^{\mathrm{T}} \\ B_2^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \delta & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

so that the matrix pair is within $O(\epsilon)$ of a pair $\hat{A}$ and $\hat{B}$ for which $\mathrm{rank}(\hat{A}\hat{B}^{\mathrm{T}}) = 1$. Our proposed algorithm tries to introduce rank deficiency into $\tilde{A}\tilde{B}^{\mathrm{T}}$ for the pair

$$\tilde{A}_{22} = \begin{bmatrix} \breve{A}_1 \mid 0 \end{bmatrix} = \left[\begin{array}{cc|ccc} \delta & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right]$$

$$\tilde{B}_2 = \begin{bmatrix} \breve{B}_1 \mid \breve{B}_2 \end{bmatrix} = \left[\begin{array}{cc|ccc} 1 & 1 & \delta & 0 & 0 \\ -\epsilon/\delta & -\epsilon & 0 & \delta & 0 \\ 0 & 0 & 0 & 0 & \delta \end{array}\right].$$

We seek $O(\epsilon)$ perturbations E_{22} and F_2 so that

$$\mathrm{rank}\left(\left(\tilde{A}_{22} + E_{22}\right)\left(\tilde{B}_2 + F_2\right)^{\mathrm{T}}\right) = 1.$$

Let

$$E_{22} = \begin{bmatrix} \breve{E}_1 & \breve{E}_2 \end{bmatrix}, \qquad F_2 = \begin{bmatrix} \breve{F}_1 & \breve{F}_2 \end{bmatrix}.$$

We assume that

$$\begin{aligned} \left(\tilde{A}_{22} + E_{22}\right)\left(\tilde{B}_2 + F_2\right)^{\mathrm{T}} &= \begin{bmatrix} \breve{A}_1 + \breve{E}_1 & \breve{E}_2 \end{bmatrix} \begin{bmatrix} \breve{B}_1^{\mathrm{T}} + \breve{F}_1^{\mathrm{T}} \\ \breve{B}_2^{\mathrm{T}} + \breve{F}_2^{\mathrm{T}} \end{bmatrix} \\ &= \breve{A}_1\breve{B}_1^{\mathrm{T}} + \breve{A}_1\breve{F}_1^{\mathrm{T}} + \breve{E}_1\breve{B}_1^{\mathrm{T}} + \breve{E}_2\breve{B}_2^{\mathrm{T}} + O(\epsilon^2) \end{aligned}$$

has rank 1. Since $\breve{B}_2 = \delta I$ the Eckhart-Young theorem implies that the only way this can happen is if

$$\breve{A}_1\breve{B}_1^{\mathrm{T}} + \breve{A}_1\breve{F}_1^{\mathrm{T}} + \breve{E}_1\breve{B}_1^{\mathrm{T}} = \begin{bmatrix} \delta & -\epsilon & 0 \\ 1 & -\epsilon & 0 \end{bmatrix} + \begin{bmatrix} \delta & 0 \\ 0 & 1 \end{bmatrix} \breve{F}_1^{\mathrm{T}} + \breve{E}_1 \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} + O(\epsilon^2)$$

has a singular value of $O(\delta\epsilon)$. If $\|\breve{E}_1\| = O(\epsilon)$ and $\|\breve{F}_1\| = O(\epsilon)$ then the first two columns of this matrix have the form

$$X = \begin{bmatrix} \delta + O(\epsilon) & -\epsilon + O(\delta\epsilon) \\ 1 + O(\epsilon) & -\epsilon + O(\epsilon) \end{bmatrix}.$$

Without the $O(\epsilon)$ terms this matrix has rank 1. It has a left singular vector associated with the smallest singular value of the form

$$u^{\mathrm{T}} = \begin{bmatrix} \frac{-1}{\sqrt{1+\delta^2}} & \frac{\delta}{\sqrt{1+\delta^2}} \end{bmatrix} + O(\epsilon).$$

We can find an expansion to estimate the smallest singular value. Theorem 3 implies that

$$\sigma_2^2(X) = \left| u^{\mathrm{T}} \begin{bmatrix} O(\epsilon) & -\epsilon + O(\delta\epsilon) \\ O(\epsilon) & -\epsilon + O(\epsilon) \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right|^2 + O(\epsilon^3) = (\epsilon + O(\delta\epsilon))^2 .$$

Ignoring the third column can only decrease σ_2 so

$$\sigma_2(\breve{A}_1\breve{B}_1^{\mathrm{T}} + \breve{A}_1\breve{F}_1^{\mathrm{T}} + \breve{E}_1\breve{B}_1^{\mathrm{T}}) \geq \sigma_2(X) \geq \epsilon - O(\delta\epsilon) \gg \delta\epsilon.$$

Thus if $\|\breve{E}_1\|$ and $\|\breve{F}_1\|$ are not much larger than ϵ then we must have $\|\breve{E}_2\| \gg \epsilon$. It can also be experimentally verified that Theorem 1 applied to $\tilde{A}_{22}$ and $\tilde{B}_2$ results in rank reducing perturbations that are much larger than ϵ. ■

5 Conclusions

We have proposed a new method for detecting near rank deficiency in the product AB^{T}. To first order its performance is provably insensitive to ill-conditioning of the matrices. It is also insensitive to the presence of nearly intersecting subspaces associated with moderately small singular values—a situation that can degrade the accuracy of subspaces estimated using the product SVD.

We have also proposed a natural method for attempting to find higher order rank deficiency based on generalizing the orthogonality properties of SVD and the Eckhart-Young rank reducing perturbations. Unfortunately, the method was shown to be inadequate in the product case. Finding an algorithm for reliably revealing the rank of a product of perturbed matrices remains an open problem.

References

[1] B. De Moor and P. Van Dooren. Generalizations of the singular value and QR decompositions. *SIAM Journal of Matrix Analysis and Applications*, 13:993–1014, 1992.

[2] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, 3nd edition, 1996.

[3] C. C. Paige. Some aspects of generalized QR factorizations. In M. G. Cox and S. J. Hammarling, editors, *Reliable Numerical Computation*, pages 71–91, Oxford, 1990. Clarendon Press.

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

[5] G. W. Stewart. A second order perturbation expansion for small singular values. *Linear Algebra and Its Applications*, 56:231–235, 1984.