

On the Minimum Rank of a Generalized Matrix Approximation Problem in the Maximum Singular Value Norm

Kin Cheong Sou and Anders Rantzer

Abstract—In this paper theoretical results regarding a generalized minimum rank matrix approximation problem in the maximum singular value norm are presented. Using the idea of projection, the considered problem can be shown to be equivalent to a classical minimum rank matrix approximation which can be solved efficiently using singular value decomposition. In addition, as long as the generalized problem is feasible, it is shown to have exactly the same optimal objective value as that of the classical problem. Certain comments and extensions of the presented theorem are included in the end of the paper.

I. INTRODUCTION

Let m, n, m_X, n_X be positive integers. Define the following three matrices and assume that

$$\begin{aligned} A \in \mathbb{R}^{m \times n}, \quad B \in \mathbb{R}^{m \times m_X}, \quad C \in \mathbb{R}^{n_X \times n} \\ m > m_X \text{ and } B \text{ has full column rank} \\ n > n_X \text{ and } C \text{ has full row rank} \end{aligned} \quad (1)$$

This paper is concerned with the following generalized minimum rank matrix approximation problem.

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \text{rank}(X) \\ \text{subject to} \quad & \|A + BXC\|_2 \triangleq \bar{\sigma}(A + BXC) < 1 \end{aligned} \quad (2)$$

This is a generalization of the following classical problem.

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \text{rank}(X) \\ \text{subject to} \quad & \|M + X\|_2 < 1 \end{aligned} \quad (3)$$

for any data matrix M , which plays the role of A in (2). In the subsequent discussion, (3) will also be referred to as the unrestricted version of (2). As it will be described in Section II-B, the classical problem in (3) can be solved efficiently using singular value decomposition (SVD) due to the theorem of Eckart-Young-Mirsky. However, it is not immediately obvious whether the generalized version in (2) is efficiently solvable or not because of the dimensions of B and C assumed in (1). This question will be answered by the main theorem in this paper.

The problem in (2) has roots in both the linear algebra and control system communities. On one hand, (2) is one of the many generalizations of (3) (e.g. [1], [2], [3], [4], [5]). On the other hand, (2) is closely related to the subject of matrix completion and the Parrott's Theorem with applications to the optimal \mathcal{H}_∞ control problem (e.g. [6], [7], [8]). Out of the many previous results, the current paper resembles [5] and [8] (Theorem 2.22 p.42) the most. The main difference between [5] and the current paper is that in the former case the norm in the optimization problem corresponding

to (2) is the Frobenius norm. The difference in norm leads to some ramifications, especially in the way how B and C increase the minimum rank of (2), as opposed to that of the unrestricted version in (3). This is to be explained in detail in Section IV-B. Regarding [8], the current paper differs from the former in the way the main theorem is proved. In this paper, the main result is obtained with the idea of projection and simple linear algebra. This provides a new perspective on the matrix approximation/completion problem. In addition, this new perspective leads to a result on a generalization of (2) which is not obvious to obtain using the result in [8]. This extension is explained in detail in Section IV-C. The rest of this paper is organized as follows. In Section II some background materials necessary to the development of the paper are described. Then in Section III the main result of this paper is presented. After that, in Section IV relevant comments and extensions to the main theorem will be discussed.

II. NOTATIONS AND BACKGROUND

A. Notations

To describe the main result, it is necessary to introduce the following SVD computable terms related to the data matrices B and C . Denote the rectangular or ‘‘economy size’’ SVD of B and C as

$$\begin{aligned} B = U_B S_B V_B^T \quad \text{s.t.} \quad & U_B \in \mathbb{R}^{m \times m_X}, \quad U_B^T U_B = I_{m_X} \\ & S_B \in \mathbb{R}^{m_X \times m_X}, \quad \text{diagonal} \\ & V_B \in \mathbb{R}^{m_X \times m_X}, \quad V_B^T V_B = I_{m_X} \end{aligned} \quad (4)$$

$$\begin{aligned} C = U_C S_C V_C^T \quad \text{s.t.} \quad & U_C \in \mathbb{R}^{n_X \times n_X}, \quad U_C^T U_C = I_{n_X} \\ & S_C \in \mathbb{R}^{n_X \times n_X}, \quad \text{diagonal} \\ & V_C \in \mathbb{R}^{n_X \times n_X}, \quad V_C^T V_C = I_{n_X} \end{aligned} \quad (5)$$

In addition, define N_B and N_C as orthonormal basis matrices for the kernels of U_B^T in (4) and V_C^T in (5), respectively.

$$\begin{aligned} N_B \in \mathbb{R}^{m \times (m - m_X)}, \quad N_B^T N_B = I, \quad U_B^T N_B = 0 \\ N_C \in \mathbb{R}^{n_X \times (n_X - n_X)}, \quad N_C^T N_C = I, \quad V_C^T N_C = 0 \end{aligned} \quad (6)$$

Notice from (4), (5), (6) that the matrices $\begin{bmatrix} N_B & U_B \end{bmatrix}$ and $\begin{bmatrix} N_C & V_C \end{bmatrix}$ are unitary (i.e. they are square, and the transposes are their inverses, respectively). Therefore, it holds that

$$\begin{aligned} N_B N_B^T + U_B U_B^T &= \begin{bmatrix} N_B & U_B \end{bmatrix} \begin{bmatrix} N_B & U_B \end{bmatrix}^T = I_m \\ N_C N_C^T + V_C V_C^T &= \begin{bmatrix} N_C & V_C \end{bmatrix} \begin{bmatrix} N_C & V_C \end{bmatrix}^T = I_n \end{aligned} \quad (7)$$

The authors are with the Department of Automatic Control, Lund University, Lund, Sweden, {kcsou, rantzer}@control.lth.se

B. Classical low rank matrix approximation via SVD

For any matrix M of rank r and an integer $k \geq 0$, the following operation is important for the solutions of the matrix approximation problems in this paper. Let the SVD of M be $M = \sum_{i=1}^r u_i \sigma_i v_i^T$, where u_i and v_i are the left and right singular vectors and $\sigma_i > 0$ are the descending singular values of M . Then the rank k truncation of M , denoted as $[M]_k$, is defined as

$$[M]_k \triangleq \begin{cases} M & k > r \\ \sum_{i=1}^k u_i \sigma_i v_i^T & 1 \leq k \leq r \\ 0 & k = 0 \end{cases} \quad (8)$$

Obviously, the rank k truncation of M has rank k if $k \leq r$.

The solution to the classical problem in (3) is provided by the theorem by Eckart-Young-Mirsky (e.g. [9]). In particular, the attainable minimum rank is the number of singular values of M which are greater than or equal to one. This number will be referred to as the *singular value excess* of M for the rest of the paper (for any matrix M). In addition, if the minimum rank is denoted as r^* , then an optimal solution to (3) can be obtained as $X^* = -[M]_{r^*}$.

C. A constrained version of (3) with known solution

For any matrix $M = [M_1 \ M_2]$ where $M_1 \in \mathbb{R}^{p \times q_1}$, $M_2 \in \mathbb{R}^{p \times q_2}$ such that M_1 has full column rank ($= q_1$), consider the problem of

$$\begin{aligned} & \underset{X_2}{\text{minimize}} \quad \text{rank} \left(\begin{bmatrix} -M_1 & X_2 \end{bmatrix} \right) \\ & \text{subject to} \quad \left\| \begin{bmatrix} M_1 & M_2 \end{bmatrix} + \begin{bmatrix} -M_1 & X_2 \end{bmatrix} \right\|_2 < 1 \end{aligned} \quad (9)$$

The above problem can be interpreted as (3) with the modification that the first columns of X are required to be the negative of the corresponding columns of M . While not explicitly mentioned, [1] provides a solution to (9). In particular, denote

$$\begin{aligned} M_1 &= U_1 S_1 V_1^T && \text{as the SVD of } M_1 \\ PM_2 &= U_1 U_1^T M_2 && \text{as } M_2 \text{ projected on } \mathcal{R}(U_1) \\ P^\perp M_2 &= M_2 - PM_2 && \text{as the complement of } PM_2 \\ r_2^* &= && \text{the singular value excess of } P^\perp M_2 \end{aligned}$$

Then the attainable minimum rank in (9) is $q_1 + r_2^*$, and an optimal solution can be obtained as

$$X_2^* = - \left(PM_2 + \left[P^\perp M_2 \right]_{r_2^*} \right) \quad (10)$$

To see the assertion, note that by [1], for any $k \geq q_1$

$$\begin{aligned} & - \left(PM_2 + \left[P^\perp M_2 \right]_{k-q_1} \right) \\ &= \underset{X_2 \mid \text{rank} \left(\begin{bmatrix} -M_1 & X_2 \end{bmatrix} \right) \leq k}{\text{argmin}} \left\| \begin{bmatrix} M_1 & M_2 \end{bmatrix} + \begin{bmatrix} -M_1 & X_2 \end{bmatrix} \right\|_2 \end{aligned} \quad (11)$$

Therefore, it holds that

$$\begin{aligned} & \min_{X_2 \mid \text{rank} \left(\begin{bmatrix} -M_1 & X_2 \end{bmatrix} \right) \leq k} \left\| \begin{bmatrix} M_1 & M_2 \end{bmatrix} + \begin{bmatrix} -M_1 & X_2 \end{bmatrix} \right\|_2 \\ &= \left\| \left(PM_2 + P^\perp M_2 \right) - \left(PM_2 + \left[P^\perp M_2 \right]_{k-q_1} \right) \right\|_2 \\ &= \sigma_{k-q_1+1} \left(P^\perp M_2 \right) \end{aligned}$$

Any k which renders the above expression less than one is a lower bound of the attainable minimum rank in (9). Therefore, the maximum lower bound, denoted as k^* , satisfies the condition that $k^* - q_1 + 1$ is the index of the largest singular value of $P^\perp M_2$ which is less than one. In other words, the minimum rank of (9) is $k^* = q_1 + r_2^*$ where r_2^* is the singular value excess of $P^\perp M_2$. Finally, substituting $k^* = q_1 + r_2^*$ into (11) gives rise to the solution in (10).

III. MAIN RESULT

Before the main result can be presented, a (known) preliminary lemma is described first to provide the conditions for the main theorem to be valid.

A. Preliminary

Lemma 1: Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times mx}$, $C \in \mathbb{R}^{nx \times n}$ satisfy the assumptions in (1). In addition, let the matrices U_B , S_B , V_B be defined in (4), U_C , S_C , V_C be defined in (5) and N_B , N_C be defined in (6). Then there exists a matrix $X \in \mathbb{R}^{m \times nx}$ such that

$$\|A + BXC\|_2 \triangleq \bar{\sigma}(A + BXC) < 1 \quad (12)$$

if and only if

$$\|N_B^T A\|_2 < 1 \quad \text{and} \quad \|AN_C\|_2 < 1 \quad (13)$$

□

Proof: The statement is a corollary of the Parrott's Theorem. See, for example, [8] (Corollary 2.24 p. 43). ■

B. Main result

Theorem 1: Let the data matrices be defined in the statement of Lemma 1. If (13) is true (i.e. (12) is feasible), then the following three statements are true.

(a) The inequality in (12) is equivalent to the following inequality with a new unknown \check{X} .

$$\|\check{A} + \check{X}\|_2 < 1 \quad (14)$$

where $\check{A} \in \mathbb{R}^{m \times nx}$ and

$$\begin{aligned} \check{A} &\triangleq (U_B^T \Delta_C U_B)^{-\frac{1}{2}} U_B^T \Delta_C A V_C (V_C^T \Delta_B V_C)^{\frac{1}{2}} \\ \Delta_B &\triangleq (I_n - A^T N_B N_B^T A)^{-1} \succ 0 \\ \Delta_C &\triangleq (I_m - A N_C N_C^T A^T)^{-1} \succ 0 \end{aligned} \quad (15)$$

The equivalence means that there is a one-to-one correspondence between the feasible solutions X in (12) and \check{X} in (14). The correspondence and its inverse are defined by

$$\begin{aligned} X &= V_B S_B^{-1} (U_B^T \Delta_C U_B)^{-\frac{1}{2}} \check{X} (V_C^T \Delta_B V_C)^{-\frac{1}{2}} S_C^{-1} U_C^T \\ \check{X} &= (U_B^T \Delta_C U_B)^{\frac{1}{2}} S_B V_B^T X U_C S_C (V_C^T \Delta_B V_C)^{\frac{1}{2}} \end{aligned} \quad (16)$$

(b) The following generalized minimum rank matrix approximation problem (a copy of (2) in Section I)

$$\begin{aligned} & \underset{X}{\text{minimize}} \quad \text{rank}(X) \\ & \text{subject to} \quad \|A + BXC\|_2 < 1 \end{aligned} \quad (17)$$

which is feasible because of (13), is equivalent to

$$\begin{aligned} & \underset{\check{X}}{\text{minimize}} \quad \text{rank}(\check{X}) \\ & \text{subject to} \quad \|\check{A} + \check{X}\|_2 < 1 \end{aligned} \quad (18)$$

in the sense that the minimizers of the two optimization problems are one-to-one correspondent, and that the minimum ranks of the two problems are the same. In addition, an optimal solution to (17) can be obtained as

$$X^* = -V_B S_B^{-1} (U_B^T \Delta_C U_B)^{-\frac{1}{2}} [\check{A}]_{\check{r}} (V_C^T \Delta_B V_C)^{-\frac{1}{2}} S_C^{-1} U_C^T \quad (19)$$

where \check{r} is the singular value excess of \check{A} (i.e. the number of singular values of \check{A} which are greater than or equal to one), and $[\check{A}]_{\check{r}}$ is the rank \check{r} truncation of \check{A} as described in (8) in Section II-B.

(c) The attainable minimum rank in (17) is the singular value excess of A . \square

Proof: Proof of (a): The general idea of the proof is that (12) will be shown, successively, to be equivalent to some intermediate inequalities until (14) is finally reached. To begin, note that because of $N_C N_C^T + V_C V_C^T = I$ from (7), inequality (12) is equivalent to

$$\|A(N_C N_C^T + V_C V_C^T) + BXC\|_2^2 < 1$$

Then by the definition of the maximum singular value, the above inequality is equivalent to

$$\begin{aligned} & \left((AV_C V_C^T + BXC) + AN_C N_C^T \right) \\ & \left((AV_C V_C^T + BXC) + AN_C N_C^T \right)^T \prec I \end{aligned}$$

Using the identities $N_C^T N_C = I$, $V_C^T N_C = 0$ and $C N_C = U_C S_C V_C^T N_C = 0$ (cf. (6), (5)), the above inequality becomes

$$\begin{aligned} (AV_C V_C^T + BXC)(AV_C V_C^T + BXC)^T & \prec I - AN_C N_C^T A^T \\ & = (\Delta_C)^{-1} \end{aligned} \quad (20)$$

where the last equality is due to (15), and Δ_C can be legitimately defined as a positive-definite matrix because of the assumption in (13). Multiplying both sides of (20) with $(\Delta_C)^{\frac{1}{2}}$ and expanding C as $C = U_C S_C V_C^T$, (20) becomes

$$\begin{aligned} & \left((\Delta_C)^{\frac{1}{2}} AV_C V_C^T + (\Delta_C)^{\frac{1}{2}} B X U_C S_C V_C^T \right) \\ & \left((\Delta_C)^{\frac{1}{2}} AV_C V_C^T + (\Delta_C)^{\frac{1}{2}} B X U_C S_C V_C^T \right)^T \prec I \end{aligned} \quad (21)$$

Using the relationship $V_C^T V_C = I$, (21) is equivalent to

$$\left\| (\Delta_C)^{\frac{1}{2}} AV_C + (\Delta_C)^{\frac{1}{2}} B X U_C S_C \right\|_2 < 1$$

and with simplifying notations, the above expression can be written as

$$\|\check{A} + \check{X}\|_2 < 1 \quad (22)$$

with

$$\begin{aligned} \check{A} & \triangleq (\Delta_C)^{\frac{1}{2}} AV_C \\ \check{B} & \triangleq (\Delta_C)^{\frac{1}{2}} B \\ \check{X} & \triangleq X U_C S_C \end{aligned} \quad (23)$$

To summarize the progress so far, with the assumption in (13), the inequality in (12) is shown to be equivalent to (22) in terms of \check{A} , \check{B} and \check{X} defined in (23). Note that the corresponding \check{C} matrix in (22) is now an identity. In addition, X and \check{X} are one-to-one correspondent because $U_C S_C$ is invertible. The next step in the proof is to apply the above idea again to arrive at an expression in which the corresponding “ B ” matrix will also be an identity. However, before this second step certain notations need to be introduced first.

Since B is assumed to have full column rank, \check{B} in (23) also has full column rank. Therefore, the SVD of \check{B} can be written as

$$\check{B} = U_{\check{B}} S_{\check{B}} V_{\check{B}}^T \quad \text{s.t.} \quad \begin{aligned} U_{\check{B}} & \in \mathbb{R}^{m \times m}, U_{\check{B}}^T U_{\check{B}} = I_{m \times m} \\ S_{\check{B}} & \in \mathbb{R}^{m \times m}, \text{ diagonal} \\ V_{\check{B}} & \in \mathbb{R}^{m \times m}, V_{\check{B}}^T V_{\check{B}} = I_{m \times m} \end{aligned} \quad (24)$$

Also, define $N_{\check{B}}$ as an orthonormal basis matrix for the kernel of $U_{\check{B}}^T$ as

$$N_{\check{B}} \in \mathbb{R}^{m \times (m-mx)}, \quad N_{\check{B}}^T N_{\check{B}} = I, \quad U_{\check{B}}^T N_{\check{B}} = 0 \quad (25)$$

It can also be verified that $[N_{\check{B}} \quad U_{\check{B}}]$ is a unitary matrix and therefore the following holds.

$$N_{\check{B}} N_{\check{B}}^T + U_{\check{B}} U_{\check{B}}^T = [N_{\check{B}} \quad U_{\check{B}}] [N_{\check{B}} \quad U_{\check{B}}]^T = I \quad (26)$$

Now the proof of the equivalence between (12) and (14) can be resumed, with the starting point being (22). From (26) it can be seen that (22) is equivalent to

$$\|(N_{\check{B}} N_{\check{B}}^T + U_{\check{B}} U_{\check{B}}^T) \check{A} + \check{B} \check{X}\|_2^2 < 1$$

Again, by the definition of the maximum singular value, the above inequality is equivalent to

$$\begin{aligned} & \left((U_{\check{B}} U_{\check{B}}^T \check{A} + \check{B} \check{X}) + N_{\check{B}} N_{\check{B}}^T \check{A} \right)^T \\ & \left((U_{\check{B}} U_{\check{B}}^T \check{A} + \check{B} \check{X}) + N_{\check{B}} N_{\check{B}}^T \check{A} \right) \prec I \end{aligned}$$

Using $N_{\check{B}}^T N_{\check{B}} = I$, $U_{\check{B}}^T N_{\check{B}} = 0$ and $\check{B}^T N_{\check{B}} = V_{\check{B}}^T S_{\check{B}}^T U_{\check{B}}^T N_{\check{B}} = 0$ (cf. (24) and (25)), the above inequality is equivalent to

$$\left(U_{\check{B}} U_{\check{B}}^T \check{A} + \check{B} \check{X} \right)^T \left(U_{\check{B}} U_{\check{B}}^T \check{A} + \check{B} \check{X} \right) \prec I - \check{A}^T N_{\check{B}} N_{\check{B}}^T \check{A} \quad (27)$$

It can be shown (in the Appendix) that, under the assumption in (13), the term $I - \check{A}^T N_{\check{B}} N_{\check{B}}^T \check{A}$ in the right-hand-side of (27) is positive-definite, and its inverse, denoted as $\Delta_{\check{B}}$ can be described by the “non-tilde” matrices as

$$\Delta_{\check{B}} \triangleq (I - \check{A}^T N_{\check{B}} N_{\check{B}}^T \check{A})^{-1} = V_C^T \Delta_B V_C \succ 0 \quad (28)$$

Then, multiplying both sides of (27) with $(\Delta_{\check{B}})^{\frac{1}{2}}$ and expanding \check{B} as $\check{B} = U_{\check{B}} S_{\check{B}} V_{\check{B}}^T$, (27) becomes

$$\begin{aligned} & \left(U_{\check{B}} U_{\check{B}}^T \check{A} (\Delta_{\check{B}})^{\frac{1}{2}} + U_{\check{B}} S_{\check{B}} V_{\check{B}}^T \check{X} (\Delta_{\check{B}})^{\frac{1}{2}} \right)^T \\ & \left(U_{\check{B}} U_{\check{B}}^T \check{A} (\Delta_{\check{B}})^{\frac{1}{2}} + U_{\check{B}} S_{\check{B}} V_{\check{B}}^T \check{X} (\Delta_{\check{B}})^{\frac{1}{2}} \right) \prec I \end{aligned} \quad (29)$$

Using the relationship $U_{\tilde{B}}^T U_{\tilde{B}} = I$, (29) is equivalent to

$$\left\| U_{\tilde{B}}^T \tilde{A} (\Delta_{\tilde{B}})^{\frac{1}{2}} + S_{\tilde{B}} V_{\tilde{B}}^T \tilde{X} (\Delta_{\tilde{B}})^{\frac{1}{2}} \right\|_2 < 1 \quad (30)$$

At this point, since both $S_{\tilde{B}} V_{\tilde{B}}^T$ and $(\Delta_{\tilde{B}})^{\frac{1}{2}}$ are invertible, it is possible to apply a change of variable $\check{X} = S_{\tilde{B}} V_{\tilde{B}}^T \tilde{X} (\Delta_{\tilde{B}})^{\frac{1}{2}}$ to arrive at an inequality in the form of (14). However, to obtain (14) with \check{A} and \check{X} represented by the original, “non-tilde” terms as in (15). The following expressions (proved in the Appendix) are needed.

$$\begin{aligned} U_{\tilde{B}} &= (\Delta_C)^{\frac{1}{2}} U_B (U_B^T \Delta_C U_B)^{-\frac{1}{2}} Q \\ S_{\tilde{B}} V_{\tilde{B}}^T &= Q^T (U_B^T \Delta_C U_B)^{\frac{1}{2}} S_B V_B^T \\ N_{\tilde{B}} &= (\Delta_C)^{-\frac{1}{2}} N_B (N_B^T (\Delta_C)^{-1} N_B)^{-\frac{1}{2}} Q_1 \end{aligned} \quad (31)$$

where Q and Q_1 are unitary matrices whose exact forms are irrelevant to the discussion in here. Using all the expressions of the “tilde” quantities in (31), (23) and (28), inequality (30) can be written as

$$\left\| Q^T (U_B^T \Delta_C U_B)^{-\frac{1}{2}} U_B^T \Delta_C A V_C (V_C^T \Delta_B V_C)^{\frac{1}{2}} + Q^T (U_B^T \Delta_C U_B)^{\frac{1}{2}} S_B V_B^T X U_C S_C (V_C^T \Delta_B V_C)^{\frac{1}{2}} \right\|_2 < 1$$

with Q being a unspecified unitary matrix. However, since the matrix 2-norm is unitarily invariant, the above inequality is equivalent to the following one without the Q matrix.

$$\left\| \underbrace{(U_B^T \Delta_C U_B)^{-\frac{1}{2}} U_B^T \Delta_C A V_C (V_C^T \Delta_B V_C)^{\frac{1}{2}}}_{=\check{A}} + \underbrace{(U_B^T \Delta_C U_B)^{\frac{1}{2}} S_B V_B^T X U_C S_C (V_C^T \Delta_B V_C)^{\frac{1}{2}}}_{=\check{X}} \right\|_2 < 1$$

This is the same as (14) with \check{A} defined in (15) and \check{X} defined in (16). Finally, the first line in (16) is true because both $(U_B^T \Delta_C U_B)^{\frac{1}{2}} S_B V_B^T$ and $U_C S_C (V_C^T \Delta_B V_C)^{\frac{1}{2}}$ are invertible.

Proof of (b): The equivalence between (12) and (14) establishes that the feasible solutions of the two optimization problems (17) and (18) are one-to-one correspondent. To complete the proof that (17) and (18) are equivalent optimization problems, it remains to show that the corresponding feasible solutions have the same objective value (i.e. rank). This is true, because in (16) both $(U_B^T \Delta_C U_B)^{\frac{1}{2}} S_B V_B^T$ and $U_C S_C (V_C^T \Delta_B V_C)^{\frac{1}{2}}$ are invertible. Therefore, (17) and (18) are equivalent. Next, since (18) is a classical problem in the form of (3) in Section II-B, its optimal rank, denoted as \check{r} , is the singular value excess of \check{A} . In addition, an optimal solution to (18) is $-\check{A}_{\check{r}}$ as the rank \check{r} truncation of \check{A} described in Section II-B. Finally, applying the relationship in (16), it can be seen that an optimal solution to (17) is

$$X^* = -V_B S_B^{-1} (U_B^T \Delta_C U_B)^{-\frac{1}{2}} [\check{A}]_{\check{r}} (V_C^T \Delta_B V_C)^{-\frac{1}{2}} S_C^{-1} U_C^T$$

This is the same as X^* in (19).

Proof of (c): As it was argued in the proof of (b), the optimal rank in (17) is the same as that of its equivalence

(18). The latter value, by the classical result in Section II-B, is the singular value excess of \check{A} . To complete the proof of item (c), it remains to show that the singular value excesses of \check{A} and A are the same. Alternatively, denote $k(M)^-$ as the number of non-positive eigenvalues of any matrix M with real eigenvalues only, then the desired statement to prove is

$$k(I - \check{A}\check{A}^T)^- = k(I - AA^T)^- \quad (32)$$

This proof is developed in two steps. The first step is

$$\begin{aligned} &k(I - \check{A}^T \check{A})^- \\ &= k(I - V_C^T A^T \Delta_C A V_C)^- \\ &= k(I - A V_C V_C^T A^T \Delta_C)^- \\ &= k(((\Delta_C)^{-1} - A V_C V_C^T A^T) \Delta_C)^- \\ &= k((I - A(N_C N_C^T + V_C V_C^T) A^T) \Delta_C)^- \\ &= k((I - AA^T) \Delta_C)^- \\ &= k((\Delta_C)^{\frac{1}{2}} (I - AA^T) (\Delta_C)^{\frac{1}{2}})^- \end{aligned}$$

In the above derivation, the first equality is due to the definition of \check{A} in (23). The second equality is due to the fact that the sets of nonzero eigenvalues of $V_C^T A^T \Delta_C A V_C$ and $A V_C V_C^T A^T \Delta_C$ are the same. The fourth equality is due to the definition of Δ_C in (15). The fifth equality is due to the identity $N_C N_C^T + V_C V_C^T = I$ in (7). The last equality is due to the fact that eigenvalues are preserved under similarity transforms. In fact, the above equalities hold even if the “number” is replaced with the “set” of non-positive eigenvalues. Next, by the Sylvester’s law of inertia (e.g. [9], p.223), $k(I - AA^T)^- = k((\Delta_C)^{\frac{1}{2}} (I - AA^T) (\Delta_C)^{\frac{1}{2}})^-$. Therefore, it has been established that

$$k(I - AA^T)^- = k((\Delta_C)^{\frac{1}{2}} (I - AA^T) (\Delta_C)^{\frac{1}{2}})^- = k(I - \check{A}^T \check{A})^- \quad (33)$$

This concludes the first step of the proof connecting A with \check{A} . The second step, connecting \check{A} with \check{A} , can be proved in similar fashions. In particular,

$$\begin{aligned} &k(I - \check{A}\check{A}^T)^- \\ &= k(I - U_{\tilde{B}}^T \check{A} \Delta_{\tilde{B}} \check{A}^T U_{\tilde{B}})^- \\ &= k(I - \check{A}^T U_{\tilde{B}} U_{\tilde{B}}^T \check{A} \Delta_{\tilde{B}})^- \\ &= k(((\Delta_{\tilde{B}})^{-1} - \check{A}^T U_{\tilde{B}} U_{\tilde{B}}^T \check{A}) \Delta_{\tilde{B}})^- \\ &= k((I - \check{A}^T (N_{\tilde{B}} N_{\tilde{B}}^T + U_{\tilde{B}} U_{\tilde{B}}^T) \check{A}) \Delta_{\tilde{B}})^- \\ &= k((I - \check{A}^T \check{A}) \Delta_{\tilde{B}})^- \\ &= k((\Delta_{\tilde{B}})^{\frac{1}{2}} (I - \check{A}^T \check{A}) (\Delta_{\tilde{B}})^{\frac{1}{2}})^- \end{aligned}$$

In the above derivation, the first equality is due to the definition of \check{A} in (15). The second equality is due to the fact that the nonzero eigenvalues of $U_{\tilde{B}}^T \check{A} \Delta_{\tilde{B}} \check{A}^T U_{\tilde{B}}$ and $\check{A}^T U_{\tilde{B}} U_{\tilde{B}}^T \check{A} \Delta_{\tilde{B}}$ are the same. The fourth equality is due to the definition of $\Delta_{\tilde{B}}$ in (28). The fifth equality is due to the identity $N_{\tilde{B}} N_{\tilde{B}}^T + U_{\tilde{B}} U_{\tilde{B}}^T = I$ in (26). The last equality is due to the fact that eigenvalues are preserved under similarity transforms. Next, again by the Sylvester’s law of inertia $k(I - \check{A}^T \check{A})^- = k((\Delta_{\tilde{B}})^{\frac{1}{2}} (I - \check{A}^T \check{A}) (\Delta_{\tilde{B}})^{\frac{1}{2}})^-$. Hence,

$$k(I - \check{A}^T \check{A})^- = k((\Delta_{\tilde{B}})^{\frac{1}{2}} (I - \check{A}^T \check{A}) (\Delta_{\tilde{B}})^{\frac{1}{2}})^- = k(I - \check{A}\check{A}^T)^- \quad (34)$$

Finally, combining (33) and (34) leads to (32). This concludes the proof. \blacksquare

IV. COMMENTS AND EXTENSIONS

A. General comments

As it was mentioned in the proof, the important consequence of the equivalence between the inequalities in (12) and (14) is that the apparently difficult optimization problem in (17) can be reduced into the unrestricted classical problem in (18), to which the optimal solution is readily available.

The presented theorem also applies to the situation of $\|A + BXC\|_2 < \gamma$ for any $\gamma > 0$ because the former is true if and only if $\|\gamma^{-1}A + B(\gamma^{-1}X)C\|_2 < 1$.

There is parallel version of the presented result. In particular, in the proof of item (a) instead of first projecting A^T on the column space of V_C (by applying the relationship $N_C N_C^T + V_C V_C^T = I$ to (12)), it is possible to project A on the column space of U_C first. This will result in different expressions of \check{A} and \check{X} . However, the new inequality and optimization problem with the new \check{A} and \check{X} will be equivalent to the ones in (14) and (18). In addition, the minimum rank result in item (c) will remain the same.

In fact, the Parrott's Theorem (e.g. [8] Theorem 2.22, p.42) also contains a complete characterization of the set of all X satisfying (12). There is even a generalized Parrott's Theorem [10] providing a characterization which is amenable to an additional rank constraint in X (in the spirit of the rank minimization problem in (17)). However, the characterizations by the (generalized) Parrott's Theorem do not lend themselves easily to the type of conclusion that is described in item (c) of the presented theorem.

B. Comparing the minimum ranks in the 2-norm and Frobenius norm cases

It is intuitive that in (17) B and C (with the assumed dimensions in (1)) should increase the attainable minimum rank, as opposed to the unrestricted case in (3). The question of how restrictive B and C can be is answered by the presented theorem as follows. Lemma 1 states that B and C can render (17) infeasible. However, once (17) is feasible, item (c) states that the attainable minimum rank of (17) is independent of B and C , and it is exactly the same as that of the unrestricted problem in (3), being the singular value excess of A . By the expression of the optimal solution in (19), it is noted that the minimum rank is also the singular value excess of \check{A} in (18). However, in most cases, a statement in terms of A would be more useful because A is part of the original problem data but \check{A} is not.

To numerically demonstrate the analysis result in the previous paragraph, and to compare the 2-norm related results in this paper with similar Frobenius norm related results in [5], an experiment will be described. Let A , B and C be defined in (1) and consider the following two families of problems, parameterized by an integer k .

$$J_2(k) \triangleq \underset{X}{\text{minimize}} \quad \|A + BXC\|_2 \quad (35)$$

$$\text{subject to} \quad \text{rank}(X) < k$$

and

$$J_F(k) \triangleq \underset{X}{\text{minimize}} \quad \|A + BXC\|_F \quad (36)$$

$$\text{subject to} \quad \text{rank}(X) < k$$

for $1 \leq k \leq \bar{k} \triangleq \min\{m_X, n_X\} + 1$. Before any result can be shown, however, the procedures for solving (35) and (36) should be described first. To numerically solve (35), the following family of problems are solved using (19) in the presented theorem.

$$K_2(\gamma) \triangleq \underset{X}{\text{minimize}} \quad \text{rank}(X) \quad (37)$$

$$\text{subject to} \quad \|\gamma^{-1}A + B(\gamma^{-1}X)C\|_2 < 1$$

where γ is from a grid between $\underline{\gamma} \triangleq \max\{\|N_B^T A\|_2, \|AN_C\|_2\}$ and $\bar{\gamma} \triangleq \|A\|_2$. Then a plot of $K_2(\gamma)$ in (37) can be obtained showing the minimum rank as a function of γ . However, this function can be inverted (i.e. the plot can be rotated) to obtain a plot of $J_2(k)$ in (35). As for (36), [5] provides a formula for the solution, and hence a plot of $J_F(k)$ can also be obtained. In the numerical experiment, two triplets of data (A_1, B_1, C_1) and (A_2, B_2, C_2) are randomly generated such that $A_1 = A_2$, $B_1 \neq B_2$, $C_1 \neq C_2$. For each triplet of data (35) and (36) are solved with $1 \leq k \leq \bar{k}$. Then the corresponding $J_2(k)$ and $J_F(k)$ plots are obtained (four plots in total). The plots with the 2-norm cases are shown in Fig. 1, while the plots with the Frobenius norm cases are shown in Fig. 2.

Fig. 1 shows the $J_2(k)$ plots for the two data sets (in circles and triangles) in the 2-norm setting. The figure numerically verifies the analysis result in the previous paragraph. In particular, regardless of the value of k (or the rank of X), the attainable approximation error is lower bounded by $\underline{\gamma} = \max\{\|N_B^T A\|_2, \|AN_C\|_2\}$. Alternatively, both $\|\gamma^{-1}N_B^T A\|_2$ and $\|\gamma^{-1}AN_C\|_2$ should be less than one as specified by Lemma 1. On the other hand, for smaller values of k (e.g. $k \leq 8$ in the circle case), the attainable approximation error is unaffected by B and C , and is the same as the (decreasingly) ordered singular value of A (the crosses). This is in accordance with item (c) in the presented theorem. Fig. 2, on the other hand, shows the $J_F(k)$ plots for the two data sets (in circles and triangles) in the Frobenius norm setting. Contrary to the 2-norm case, the attainable approximation error in Fig. 2 depends on B and C except when $k = 1$ (i.e. X is a zero matrix), and is different from the error in the corresponding unrestricted case (in which B and C are identity matrices). It is not known if the approximation error has any simple relationship with A , B , and C as in the 2-norm case, and no such result is reported in [5].

C. A constrained version of (17)

The presented theorem also provides the solution to the following constrained version of (17).

$$\underset{X_2}{\text{minimize}} \quad \text{rank} \left(\begin{bmatrix} M_1 & X_2 \end{bmatrix} \right) \quad (38)$$

$$\text{subject to} \quad \|A + B \begin{bmatrix} M_1 & X_2 \end{bmatrix} C\|_2 < 1$$

where $M_1 \in \mathbb{R}^{m \times n_{X1}}$, $X_2 \in \mathbb{R}^{m \times n_{X2}}$ and $n_X = n_{X1} + n_{X2}$. The only difference between (38) and (17) is that in the former problem the first columns of the decision variable “ X ” are constrained to be M_1 , a pre-specified matrix. In the following discussion it is assumed that the data (A, B, C, M_1) are chosen so that (38) is feasible.

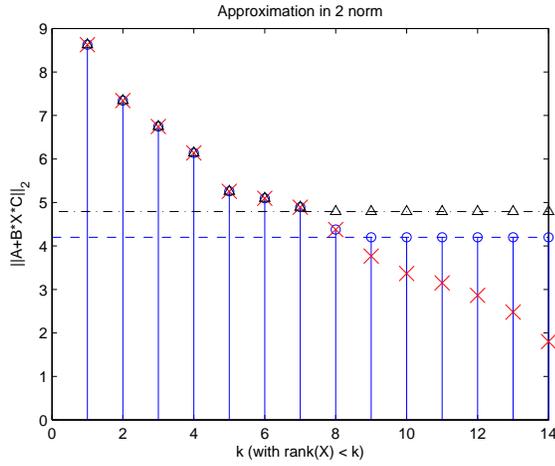


Fig. 1. The optimal 2-norm approximation errors $J_2(k)$ as functions of k in (35) for two sets of problem data. Blue circles (with vertical lines): J_2 with data (A_1, B_1, C_1) . Blue horizontal dashed line: lowest attainable J_2 with data (A_1, B_1, C_1) , regardless of the rank of X . Black triangles: J_2 with data (A_2, B_2, C_2) . Black horizontal dash-dotted line: lowest attainable J_2 with data (A_2, B_2, C_2) , regardless of the rank of X . Red crosses: ordered singular values of $A_1 = A_2$. The ordered singular values of A_i are the errors in the unrestricted case in which B_i and C_i are identity matrices. In the 2-norm case, the attainable error for smaller k (e.g. $k \leq 8$ in the circle case) is independent of B and C and is the same as that in the unrestricted case. The only effect of B and C is the cutoff in the attainable error for larger k (e.g. $k > 8$ in the circle case).

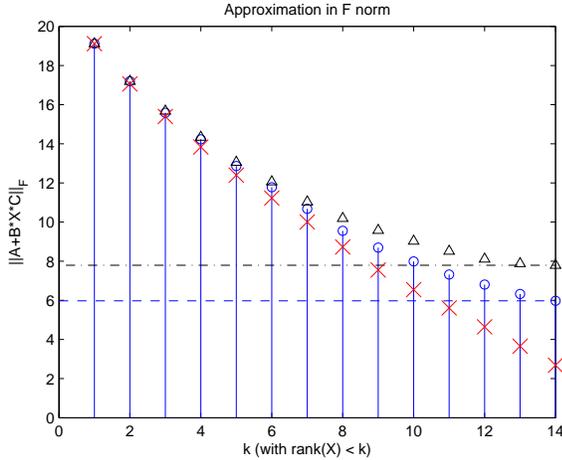


Fig. 2. The optimal Frobenius norm approximation errors $J_F(k)$ as functions of k in (36) for two sets of problem data. Blue circles (with vertical lines): J_F with data (A_1, B_1, C_1) . Blue horizontal dashed line: lowest attainable J_F with data (A_1, B_1, C_1) , regardless of the rank of X . Black triangles: J_F with data (A_2, B_2, C_2) . Black horizontal dash-dotted line: lowest attainable J_F with data (A_2, B_2, C_2) , regardless of the rank of X . Red crosses: the squared-roots of the cumulative sum of squares of the ordered singular values of $A_1 = A_2$. These quantities of A_i are the errors in the unrestricted case in which B_i and C_i are identity matrices. In the Frobenius norm case, the effect of B and C is not as simple as in the 2-norm case. B and C affect the attainable error for all values of k except $k = 1$ (when $X = 0$).

The approach to solve (38) is to use the equivalence between (12) and (14) to reduce (38) into (9) in Section II-C. Then the formula for the optimal solution in (10) can be applied. To begin, first partition C according to the dimensions of M_1 and X_2 as

$$C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}, \quad \text{with } C_1 \in \mathbb{R}^{n \times 1 \times n}, \quad C_2 \in \mathbb{R}^{n \times 2 \times n}$$

Then (38) becomes

$$\begin{aligned} & \underset{X_2}{\text{minimize}} \quad \text{rank} \left(\begin{bmatrix} M_1 & X_2 \end{bmatrix} \right) \\ & \text{subject to} \quad \|(A + BM_1C_1) + BX_2C_2\|_2 < 1 \end{aligned}$$

The above problem can be re-written as

$$\begin{aligned} & \underset{\tilde{X}}{\text{minimize}} \quad \text{rank} \left(\begin{bmatrix} M_1 & \tilde{X} \end{bmatrix} \right) \\ & \text{subject to} \quad \|\tilde{A} + \tilde{B}\tilde{X}\tilde{C}\|_2 < 1 \end{aligned} \quad (39)$$

with $\tilde{A} = A + BM_1C_1$, $\tilde{B} = B$, $\tilde{C} = C_2$ and $\tilde{X} = X_2$. The constraint in (39) has the same form as the inequality in (12). Under the feasibility assumption, this constraint is equivalent to (14) as specified by item (a) in the presented theorem. Therefore, the problem in (39) is equivalent to

$$\begin{aligned} & \underset{\check{X}}{\text{minimize}} \quad \text{rank} \left(\begin{bmatrix} M_1 & P_L\check{X}P_R \end{bmatrix} \right) \\ & \text{subject to} \quad \|\check{A} + \check{X}\|_2 < 1 \end{aligned} \quad (40)$$

with \check{A} , \check{X} , P_L , P_R , analogous to the case in (15), defined as

$$\begin{aligned} \check{A} & \triangleq (U_B^T \Delta_C U_B)^{-\frac{1}{2}} U_B^T \Delta_C \tilde{A} V_C \left(V_C^T \Delta_B V_C \right)^{\frac{1}{2}} \\ P_L & \triangleq V_B S_B^{-1} (U_B^T \Delta_C U_B)^{-\frac{1}{2}} \\ P_R & \triangleq (V_C^T \Delta_B V_C)^{-\frac{1}{2}} S_C^{-1} U_C^T \\ \check{X} & \triangleq (P_L)^{-1} \tilde{X} (P_R)^{-1} \end{aligned} \quad (41)$$

In the above expressions, the terms such as U_B , V_C and Δ_B are defined by the “non-tilde” formulas in (4), (5), (6) and (15), with the rule that the “tildes” are ignored. Since multiplying invertible matrices does not change the rank of the matrix in the objective function, the problem in (40) is equivalent to

$$\begin{aligned} & \underset{\check{X}}{\text{minimize}} \quad \text{rank} \left((P_L)^{-1} \begin{bmatrix} M_1 & P_L\check{X}P_R \\ 0 & (P_R)^{-1} \end{bmatrix} \right) \\ & \text{subject to} \quad \|\check{A} + \check{X}\|_2 < 1 \end{aligned}$$

which is also equivalent to

$$\begin{aligned} & \underset{\check{X}}{\text{minimize}} \quad \text{rank} \left(\begin{bmatrix} (P_L)^{-1} M_1 & \check{X} \end{bmatrix} \right) \\ & \text{subject to} \quad \left\| \begin{bmatrix} -(P_L)^{-1} M_1 & \check{A} \\ (P_L)^{-1} M_1 & \check{X} \end{bmatrix} \right\|_2 < 1 \end{aligned}$$

This problem is in the same form as (9), and hence the solution expression in (10) can be applied. Once a solution \check{X}^* is found, the expression from (41) can be used to find an optimal solution to (38) as $X_2^* = P_L \check{X}^* P_R$.

D. System theoretic motivating examples

The results in this paper have system theoretic motivations. In [11], [12] the following scenario is considered. A to-be-designed linear time-invariant system is described by its state-space matrices $A(L)$, $B(L)$, $C(L)$ and $D(L)$ which are affine functions of the decision matrix L . Particular to the scenario, the system matrices can be organized into

$$\begin{bmatrix} A(L) & B(L) \\ C(L) & D(L) \end{bmatrix} = F_0 + G_0 L H_0 \quad (42)$$

where F_0 , G_0 and H_0 can be computed from the data. Typically, G_0 has full column rank and H_0 has full row rank (cf. (1)). For performance and stability, it is desirable that the

system matrices satisfy the following (bounded-real lemma type) inequality.

$$\begin{bmatrix} A(L) & B(L) \\ C(L) & D(L) \end{bmatrix}^T \begin{bmatrix} X & 0 \\ 0 & \frac{1}{\gamma}I \end{bmatrix} \begin{bmatrix} A(L) & B(L) \\ C(L) & D(L) \end{bmatrix} \prec \begin{bmatrix} X & 0 \\ 0 & \gamma I \end{bmatrix} \quad (43)$$

where $\gamma > 0$ is some pre-specified scalar constant and $X \succ 0$ is, for the moment, a given matrix. Squared-root factorizing γ and X and using (42), the inequality in (43) can be re-written as

$$\|F + GLH\|_2 < 1 \quad (44)$$

with

$$\begin{aligned} F &\triangleq \begin{bmatrix} X^{\frac{1}{2}} & 0 \\ 0 & \frac{1}{\sqrt{\gamma}}I \end{bmatrix} F_0 \begin{bmatrix} X^{-\frac{1}{2}} & 0 \\ 0 & \frac{1}{\sqrt{\gamma}}I \end{bmatrix} \\ G &\triangleq \begin{bmatrix} X^{\frac{1}{2}} & 0 \\ 0 & \frac{1}{\sqrt{\gamma}}I \end{bmatrix} G_0 \\ H &\triangleq H_0 \begin{bmatrix} X^{-\frac{1}{2}} & 0 \\ 0 & \frac{1}{\sqrt{\gamma}}I \end{bmatrix} \end{aligned} \quad (45)$$

Note that (44) is in the same form as (12). In addition, to reduce the complexity of the system it is also desirable to have the rank of L minimized. Therefore, the system theoretic problem of finding a minimum rank L satisfying (44) is the same as problem (17). Item (b) in the presented theorem provides an optimal solution to the problem.

If $X \succ 0$ is also a decision variable, then the above problem becomes the following. Find $X \succ 0$ such that the minimum rank of L satisfying (44) is minimized. A convenient problem description is provided by item (c) in the presented theorem, which states that the singular value excess of F in (45) is the objective to be minimized. Using the expression of F in (45), it can be verified [12] that the singular value excess of F is the minimum objective value of the following problem.

$$\begin{aligned} &\underset{X,Y}{\text{minimize}} \quad \text{rank}(Y) \\ &\text{subject to} \quad Y \succeq F_0^T \begin{bmatrix} X & 0 \\ 0 & \frac{1}{\gamma}I \end{bmatrix} F_0 - \begin{bmatrix} X & 0 \\ 0 & \gamma I \end{bmatrix} \\ &\quad Y \succ 0 \end{aligned} \quad (46)$$

While the above optimization is a difficult rank minimization problem (in particular, not the type of (17)), a common heuristics to obtain a (usually good) sub-optimal solution is to solve a modified version of (46) in which the objective function $\text{rank}(Y)$ is replaced by $\text{trace}(Y)$. The modification is a semidefinite optimization problem, which can be solved efficiently using interior-point methods [13]. Once again, it is emphasized that without item (c), neither (46) nor its semidefinite modification could be formulated.

V. CONCLUSION

Using simple linear algebra and projection ideas, it has been shown that, under feasibility assumption, the 2-norm related inequality in (12) is equivalent to its ‘‘classical’’ version in (14). As a corollary of the above equivalence, the generalized minimum rank matrix approximation problem in (17) has also been shown to be equivalent to the classical problem in (18). Hence, an efficient solution procedure

becomes available for (17). In addition, it has been shown that the optimal objective value of (17) degrades (relative to the unrestricted problem in (3)) in a particularly simple way with respect to the problem data – the minimum rank is the same as that of the unrestricted problem except that there is a cutoff type lower bound. No analogous result is known in a similar Frobenius norm case. Finally, the equivalence between (12) and (14) also enables the efficient solution procedure for the more general optimization problem in (38), by reducing it to a case solvable with the result in [1].

ACKNOWLEDGEMENT

The authors would like to thank Prof. Alexandre Megretski and Dr. Ahmed El-Shaer for pointing out the connection of presented results with the Parrott’s Theorem.

APPENDIX

A. Proof of the expression in (28)

The outline of the proof is as follows. First it is shown that $I - \tilde{A}^T N_B N_B^T \tilde{A}$ is invertible. Then from its inverse, it can be seen that $I - \tilde{A}^T N_B N_B^T \tilde{A}$ is positive-definite. The invertibility is shown by the following arguments.

$$\begin{aligned} &\det(I - \tilde{A}^T N_B N_B^T \tilde{A}) \\ &= \det\left(I - V_C^T A^T N_B (N_B^T (\Delta_C)^{-1} N_B)^{-1} N_B^T A V_C\right) \\ &= \det\left(I - (N_B^T (\Delta_C)^{-1} N_B)^{-1} N_B^T A V_C V_C^T A^T N_B\right) \\ &= \det\left((N_B^T (\Delta_C)^{-1} N_B)^{-1}\right) \cdot \\ &\quad \det(N_B^T ((\Delta_C)^{-1} - A V_C V_C^T A^T) N_B) \\ &= \det\left((N_B^T (\Delta_C)^{-1} N_B)^{-1}\right) \cdot \\ &\quad \det(N_B^T (I - A N_C N_C^T A^T - A V_C V_C^T A^T) N_B) \\ &= \det\left((N_B^T (\Delta_C)^{-1} N_B)^{-1}\right) \det(I - N_B^T A A^T N_B) \\ &= \det\left((N_B^T (\Delta_C)^{-1} N_B)^{-1}\right) \det(I - A^T N_B N_B^T A) \\ &= \det\left((N_B^T (\Delta_C)^{-1} N_B)^{-1}\right) \det((\Delta_B)^{-1}) \\ &> 0 \end{aligned}$$

In the above derivation, the first equality is due to the definition of \tilde{A} in (23) and the expression of N_B in (31). The second equality is due to the fact that $\det(I - XY) = \det(I - YX)$ for any two matrices X and Y of suitable dimensions. The third equality is because $\det(XY) = \det(X)\det(Y)$ for any two square matrices X and Y . The fourth equality is due to the definition of Δ_C in (15). The fifth equality is due to (7). The last equality is by the definition of Δ_B in (15). Finally, the last inequality is true because by (13) $\Delta_B \succ 0$, $\Delta_C \succ 0$ and by definition N_B has full rank. Altogether, it is established that $I - \tilde{A}^T N_B N_B^T \tilde{A}$ in the right-hand-side of (27) is invertible.

Next, $(I - \tilde{A}^T N_B N_B^T \tilde{A})^{-1}$ will be shown to be $V_C^T \Delta_B V_C$,

as in (28). The argument is as follows.

$$\begin{aligned}
 & (I - \tilde{A}^T N_{\tilde{B}} N_{\tilde{B}}^T \tilde{A})^{-1} \\
 &= \left(I - V_C^T A^T N_B (N_B^T (\Delta_C)^{-1} N_B)^{-1} N_B^T A V_C \right)^{-1} \\
 &= I - V_C^T A^T N_B (N_B^T (A V_C V_C^T A^T - (\Delta_C)^{-1} N_B)^{-1} N_B^T A V_C \\
 &= I - V_C^T A^T N_B (N_B^T (A A^T - I) N_B)^{-1} N_B^T A V_C \\
 &= V_C^T \left(I - A^T N_B (N_B^T A A^T N_B - I)^{-1} N_B^T A \right) V_C \\
 &= V_C^T (I - A^T N_B N_B^T A)^{-1} V_C \\
 &= V_C^T \Delta_B V_C
 \end{aligned}$$

In the above derivation, the first equality is due to the definition of \tilde{A} in (23) and the expression of $N_{\tilde{B}}$ in (31). The second equality is due to the matrix inversion lemma [14] $(R + PSQ)^{-1} = R^{-1} - R^{-1}P(S^{-1} + QR^{-1}P)^{-1}QR^{-1}$, with $R = I$, $P = V_C^T A^T N_B$, $S = -(N_B^T (\Delta_C)^{-1} N_B)^{-1}$ and $Q = N_B^T A V_C$. The third equality is due to the definition of Δ_C in (15) and the identity $V_C V_C^T + N_C N_C^T = I$ in (7). The fourth equality is by the fact that $V_C^T V_C = I$ and $N_B^T N_B = I$. The fifth equality is again due to the matrix inversion lemma. Finally, the last equality is due to the definition of Δ_B in (15). Since by (13) $\Delta_B \succ 0$ and V_C has full column rank, $(I - \tilde{A}^T N_{\tilde{B}} N_{\tilde{B}}^T \tilde{A})^{-1} = V_C^T \Delta_B V_C \succ 0$ is shown. ■

B. Proof of the expressions in (31)

To show the first line of (31), notice from (24), (23) and (4) that the SVD of \tilde{B} can be written as

$$\tilde{B} = U_{\tilde{B}} S_{\tilde{B}} V_{\tilde{B}}^T = (\Delta_C)^{\frac{1}{2}} U_B S_B V_B^T = (\Delta_C)^{\frac{1}{2}} B \quad (47)$$

Since $S_{\tilde{B}} V_{\tilde{B}}^T$ is invertible, the second equality above implies that $U_{\tilde{B}}$ has the form

$$U_{\tilde{B}} = (\Delta_C)^{\frac{1}{2}} U_B P \quad (48)$$

with P being an invertible matrix. By the definition of SVD, $U_{\tilde{B}}$ is an orthonormal matrix, hence it holds that

$$U_{\tilde{B}}^T U_{\tilde{B}} = P^T U_B^T \Delta_C U_B P = I$$

Since $(U_B^T \Delta_C U_B)^{\frac{1}{2}} P$ is a square matrix, the above equality implies that there exists a unitary matrix Q such that

$$P = (U_B^T \Delta_C U_B)^{-\frac{1}{2}} Q$$

Substituting the above expression into (48), $U_{\tilde{B}}$ becomes

$$U_{\tilde{B}} = (\Delta_C)^{\frac{1}{2}} U_B (U_B^T \Delta_C U_B)^{-\frac{1}{2}} Q \quad (49)$$

This is the same expression as the first line in (31).

To show the second line of (31), substitute the expression of $U_{\tilde{B}}$ in (49) into (47), then the second equality implies that

$$\left((\Delta_C)^{\frac{1}{2}} U_B \right) \left((U_B^T \Delta_C U_B)^{-\frac{1}{2}} Q S_{\tilde{B}} V_{\tilde{B}}^T - S_B V_B^T \right) = 0$$

Since $(\Delta_C)^{\frac{1}{2}} U_B$ has full column rank, the above equality implies that

$$(U_B^T \Delta_C U_B)^{-\frac{1}{2}} Q S_{\tilde{B}} V_{\tilde{B}}^T - S_B V_B^T = 0$$

Alternatively, it holds that

$$S_{\tilde{B}} V_{\tilde{B}}^T = Q^T (U_B^T \Delta_C U_B)^{\frac{1}{2}} S_B V_B^T$$

This is the same as the second line in (31).

Finally, to show the third line of (31), notice from (25) and (49) that

$$U_{\tilde{B}}^T N_{\tilde{B}} = Q^T (U_{\tilde{B}}^T \Delta_C U_{\tilde{B}})^{-\frac{1}{2}} U_B^T (\Delta_C)^{\frac{1}{2}} N_{\tilde{B}} = 0$$

The fact that $Q^T (U_{\tilde{B}}^T \Delta_C U_{\tilde{B}})^{-\frac{1}{2}}$ is invertible implies that

$$U_B^T (\Delta_C)^{\frac{1}{2}} N_{\tilde{B}} = 0$$

This means that $(\Delta_C)^{\frac{1}{2}} N_{\tilde{B}}$ is in the kernel of U_B^T , which is characterized by its basis matrix N_B (cf. (6)). Hence, the above equality implies that there exists a square matrix Y (Y is square because N_B and $N_{\tilde{B}}$ have the same dimension) such that

$$N_{\tilde{B}} = (\Delta_C)^{-\frac{1}{2}} N_B Y \quad (50)$$

Also, by definition $N_{\tilde{B}}$ is an orthonormal matrix. Therefore,

$$N_{\tilde{B}}^T N_{\tilde{B}} = Y^T N_B^T (\Delta_C)^{-1} N_B Y = I$$

Since $(N_B^T (\Delta_C)^{-1} N_B)^{\frac{1}{2}}$ is a square matrix, the above identity implies that there exists a unitary matrix Q_1 such that

$$Y = (N_B^T (\Delta_C)^{-1} N_B)^{-\frac{1}{2}} Q_1$$

Substituting the above expression of Y into (50) yields

$$N_{\tilde{B}} = (\Delta_C)^{-\frac{1}{2}} N_B (N_B^T (\Delta_C)^{-1} N_B)^{-\frac{1}{2}} Q_1$$

This is the same expression as the third line in (31). ■

REFERENCES

- [1] G. Golub, A. Hoffman, and G. Stewart, "A Generalization of the Eckart-Young-Mirsky Matrix Approximation Theorem," *Linear Algebra and its Applications*, vol. 88/89, pp. 317–327, 1987.
- [2] J. Demmel, "The smallest perturbation of a submatrix which lowers the rank and constrained total least squares problems," *SIAM Journal on Numerical Analysis*, vol. 24, no. 1, p. 206, February 1987.
- [3] H. Zha, "A numerical algorithm for computing the restricted singular value decomposition of matrix tripiets," *Linear Algebra and its Applications*, vol. 168, pp. 1–25, 1992.
- [4] Y. Nievergelt, "Schmidt-Mirsky Matrix Approximation With Linearly Constrained Singular Values," *Linear Algebra and its Applications*, vol. 261, pp. 207–219, 1997.
- [5] S. Friedland and A. Torokhti, "Generalized rank-constrained matrix approximations," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 2, pp. 656–659, 2006.
- [6] A. Packard, K. Zhou, P. Pandey, J. Leonhardson, and G. Balas, "Optimal, constant i/o similarity scaling for full-information and state feedback control problems," *Systems and Control Letters*, vol. 19, pp. 271–280, 1992.
- [7] C. Beck, J. Doyle, and K. Glover, "Model reduction of multidimensional and uncertain systems," *IEEE Transactions on Automatic Control*, vol. 41, no. 10, 1996.
- [8] K. Zhou, J. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice Hall, 1996.
- [9] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.
- [10] A. Megretski, "MIT Lecture notes, 6.245 Multivariable Control Systems," available from <http://web.mit.edu/6.245/www/>.
- [11] K.C. Sou and A. Rantzer, "A singular value decomposition based closed loop stability preserving controller reduction method," in *American Control Conference*, June 2010.
- [12] K.C. Sou, "Controller Reduction with Closed Loop Error Guarantee," Submitted to the Conference on Decision and Control, December 2010.
- [13] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.
- [14] N. Higham, *Accuracy and Stability of Numerical Algorithms*. SIAM, 2002.