On a generalized matrix approximation problem in the spectral norm

Sou, Kin Cheong; Rantzer, Anders

Published in:
Linear Algebra and Its Applications

DOI:
10.1016/j.laa.2011.10.009

2012

On the Generalized Matrix Approximation Problems in the Spectral Norm

Kin Cheong Sou*, Anders Rantzer

ACCESS Linnaeus Center and the Automatic Control Lab, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, 10044, Sweden. Tel: +4687907427

LCCC Linnaeus Center and the Department of Automatic Control, Lund University, Lund, 22100, Sweden

Abstract

In this paper theoretical results regarding a generalized minimum rank matrix approximation problem in the spectral norm are presented. An alternative solution expression for the generalized matrix approximation problem is obtained. This alternative expression provides a simple characterization of the achievable minimum rank, which is shown to be the same as the optimal objective value of the classical problem considered by Eckart-Young-Schmidt-Mirsky, as long as the generalized problem is feasible. In addition, this paper provides a result on a constrained version of the matrix approximation problem, establishing that the later problem is solvable via singular value decomposition.

Keywords: Matrix approximation; Rank minimization; Singular value decomposition

1. Introduction

This paper considers the following generalized minimum rank matrix approximation problem:

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

☆This research was completed during the first author's appointment at Lund University.

*Corresponding author.

Email addresses: kin.cheong.sou@ee.kth.se (Kin Cheong Sou), rantzer@control.lth.se (Anders Rantzer)

Here the data matrices are $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times m_X}$, and $C \in \mathbb{R}^{n_X\times n}$. The symbol $\|\cdot\|_2$ denotes the spectral norm of a matrix (i.e., the maximum singular value).

Assumption 1.1. *In this paper, it is assumed that $m > m_X$ and B has full column rank. In addition, $n > n_X$ and C has full row rank.*

Remark 1.1. *The assumed dimensions and ranks on B and C ensure that (1) cannot be trivially reduced to the classical problem to be described in (2).*

The problem in (1) is a generalization of the following classical problem:

$$\begin{array}{ll} \underset{X}{\text{minimize}} & \text{rank}(X) \\ \text{subject to} & \|M+X\|_2 < 1 \end{array} \tag{2}$$

for any data matrix M, which plays the role of A in (1). The classical problem in (2) can be solved efficiently using singular value decomposition (SVD). In addition, the minimum rank in (2) can easily be characterized using the singular values of M. Though less well-known, (1) can in fact be solved via SVD using matrix dilation/Parrott's Lemma results (e.g. [1, 2, 3]). However, to the authors' best knowledge, no simple characterization of the minimum rank in (1) in terms of problem data A, B and C is known. This characterization is based on an alternative solution expression for (1), which cannot be found in [1, 2, 3]. In addition, this paper provides an SVD based solution to a constrained version of (1). This is also not available in [1, 2, 3].

There have been many efforts for the generalizations of (2) (e.g. [4, 5, 6, 7, 8]). However, none of these results apply to problem (1) considered in this paper. The most related result is [8], which considers a variant of (1) with the constraint being $\|A+BXC\|_F < 1$ (i.e., the Frobenius norm). However, this paper is fundamentally different from [8]. In particular, (1) is not a special case of the problem in [8] or vice versa. Moreover, the result and proof technique in [8] do not apply to the problem considered in this paper. Most importantly, none of the previous work, including [8], provide any simple characterization of the achievable minimum rank analogous to the main result of this paper.

In summary, this paper contains the following contributions which, to the authors' best knowledge, have not been published:

1. An alternative solution expression for (1).
2. A simple characterization of the achievable minimum rank in (1).
3. An SVD based solution procedure for a constrained version of (1).

The rest of this paper is organized as follows. In Section 2 some background material and notations necessary to the development of the paper are described. In Section 3 the main result concerning the simple characterization of the minimum rank of (1) is presented. In Section 4 the SVD based solution procedure for a constrained version of (1) is described. Finally, conclusions are made in Section 5.

2. Background

2.1. Definitions of 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 SVD of B and C as

$$B = \begin{bmatrix} U_B & N_B \end{bmatrix} \begin{bmatrix} S_B \\ 0 \end{bmatrix} {V_B}^T = U_B S_B {V_B}^T$$

$$\begin{aligned} \text{such that} \quad & U_B \in \mathbb{R}^{m \times m_X}, && {U_B}^T U_B = I_{m_X} \\ & N_B \in \mathbb{R}^{m \times (m-m_X)}, && {N_B}^T N_B = I_{m-m_X} \\ & S_B \in \mathbb{R}^{m_X \times m_X}, && \text{diagonal and positive definite} \\ & V_B \in \mathbb{R}^{m_X \times m_X}, && {V_B}^T V_B = I_{m_X}, \end{aligned} \tag{3}$$

$$C = U_C \begin{bmatrix} S_C \\ 0 \end{bmatrix} \begin{bmatrix} V_C & N_C \end{bmatrix}^T = U_C S_C {V_C}^T$$

$$\begin{aligned} \text{such that} \quad & U_C \in \mathbb{R}^{n_X \times n_X}, && {U_C}^T U_C = I_{n_X} \\ & S_C \in \mathbb{R}^{n_X \times n_X}, && \text{diagonal and positive definite} \\ & V_C \in \mathbb{R}^{n \times n_X}, && {V_C}^T V_C = I_{n_X} \\ & N_C \in \mathbb{R}^{n \times (n-n_X)}, && {N_C}^T N_C = I_{n-n_X}. \end{aligned} \tag{4}$$

Also from the SVD, the matrices $\begin{bmatrix} N_B & U_B \end{bmatrix}$ and $\begin{bmatrix} N_C & V_C \end{bmatrix}$ are orthogonal. Hence,

$$\begin{aligned} & {U_B}^T N_B = 0 \\ & {V_C}^T N_C = 0 \\ & N_B {N_B}^T + U_B {U_B}^T = \begin{bmatrix} N_B & U_B \end{bmatrix} \begin{bmatrix} N_B & U_B \end{bmatrix}^T = I_m \\ & N_C {N_C}^T + V_C {V_C}^T = \begin{bmatrix} N_C & V_C \end{bmatrix} \begin{bmatrix} N_C & V_C \end{bmatrix}^T = I_n. \end{aligned} \tag{5}$$

2.2. Classical minimum 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 non-increasing singular values of M. Then the rank k truncation of M, denoted as $[M]_k$, is defined as

$$[M]_k \triangleq \begin{cases} M & k > r \\ \sum_{i=1}^{k} u_i \sigma_i {v_i}^T & 1 \leq k \leq r \\ 0 & k = 0. \end{cases} \tag{6}$$

The classical problem in (2) can be written as

$$\begin{array}{lll} \underset{k \in \mathbb{Z}^+}{\text{minimize}} & & k \\ \text{subject to} & \underset{X}{\min} & \|M+X\|_2 \\ & \text{subject to} & \text{rank}(X) \leq k \end{array} < 1 \iff \begin{array}{ll} \underset{k \in \mathbb{Z}^+}{\text{minimize}} & k \\ \text{subject to} & \sigma_{k+1}(M) < 1, \end{array} \tag{7}$$

where $\sigma_i(M) > 0$ for $i = 1, 2, \ldots$ are the non-increasing singular values of M and the equivalence above is due to the theorem by Eckart-Young-Schmidt-Mirsky (e.g. [9]). Therefore, the minimum value of k in (7) (i.e., the minimum rank in (2)) is the number of singular values of M which are greater than or equal to one. In subsequent, this number will be referred to as the ***singular value excess*** of M, and denoted as $\text{sve}(M)$. That is,

$$\text{sve}(M) \triangleq \begin{cases} k & \text{such that} \quad \sigma_1(M) \geq \ldots \geq \sigma_k(M) \geq 1 > \sigma_{k+1}(M) \geq \ldots \\ 0 & \text{if} \quad 1 > \sigma_1(M) \\ \text{rank}(M) & \text{if} \quad \sigma_{\text{rank}(M)}(M) \geq 1. \end{cases} \tag{8}$$

Note that the definition of singular value excess in (8) also applies to matrices other than M considered here. Finally, by the theorem by Eckart-Young-Schmidt-Mirksy, an optimal solution to (2) can be obtained as $X^\star = -[M]_{\mathrm{sve}(M)}$.

3. Simple Characterization of Minimum Rank

This section describes the main result of the paper, providing a simple characterization of the minimum rank of (1). Before the main result is presented, several preliminary results should be described first.

3.1. Preliminary Results: A new equivalent constraint of (1)

The first preliminary result, stated without proof, is known as the Parrott's Lemma (e.g. [1], p.43). It provides the sufficient and necessary conditions for the generalized minimum rank matrix approximation problem in (1) to be feasible.

Proposition 3.1. *Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times m_X}$, $C \in \mathbb{R}^{n_X\times n}$ satisfy assumption 1.1. In addition, let the matrices U_B, N_B, S_B, V_B be defined in (3) and U_C, S_C, V_C, N_C be defined in (4). Then there exists a matrix $X \in \mathbb{R}^{m_X\times n_X}$ such that*

$$\|A+BXC\|_2 \triangleq \sigma_1(A+BXC) < 1 \tag{9}$$

if and only if

$$\left\|N_B{}^T A\right\|_2 < 1 \quad \text{and} \quad \|AN_C\|_2 < 1. \tag{10}$$

Remark 3.1. *If (10) holds, then the following two symmetric positive definite matrices can be defined:*

$$\begin{aligned} \Delta_B &\triangleq \left(I_n - A^T N_B N_B{}^T A\right)^{-1} \succ 0 \\ \Delta_C &\triangleq \left(I_m - AN_C N_C{}^T A^T\right)^{-1} \succ 0. \end{aligned} \tag{11}$$

Δ_B and Δ_C will be used in the subsequent discussions.

The second preliminary result is an equivalent expression of the generalized Parrott's Lemma (e.g. [1, 3]). The expression to be presented is new, and it is required to prove the main theorem in Section 3.

Proposition 3.2. *Let the data matrices be defined in the statement of Proposition 3.1. If (10) is true (i.e., (9) is feasible), then the inequality in (9) is equivalent to the following inequality with a new unknown $\check{X}$:*

$$\left\|\check{A}+\check{X}\right\|_2 < 1 \tag{12}$$

where $\check{A} \in \mathbb{R}^{m_X \times n_X}$, and is defined as

$$\check{A} \triangleq \left(U_B{}^T \Delta_C U_B\right)^{-\frac{1}{2}} U_B{}^T \Delta_C A V_C \left(V_C{}^T \Delta_B V_C\right)^{\frac{1}{2}}, \tag{13}$$

where Δ_B and Δ_C are defined in (11). The equivalence means that there is a one-to-one correspondence between the feasible solutions X in (9) and $\check{X}$ in (12). The correspondence and its inverse are defined by

$$\begin{array}{rcl} X & = & V_B S_B{}^{-1} \left(U_B{}^T \Delta_C U_B\right)^{-\frac{1}{2}} \check{X} \left(V_C{}^T \Delta_B V_C\right)^{-\frac{1}{2}} S_C{}^{-1} U_C{}^T \\ \check{X} & = & \left(U_B{}^T \Delta_C U_B\right)^{\frac{1}{2}} S_B V_B{}^T X U_C S_C \left(V_C{}^T \Delta_B V_C\right)^{\frac{1}{2}}. \end{array} \tag{14}$$

PROOF. See Appendix. ■

Remark 3.2. *Many alternative forms of (12) exist (e.g. Corollary 2.24 of [1] (p. 43)). However, the proof development of the main theorem in Section 3 requires expressions (12) and (13). The authors are not aware of any straightforward approach to arrive at the conclusion in the main theorem using any expression other than (12) and (13). Moreover, it is not known if there is any simple transformation between the alternative expressions and (12) and (13), other than the fact that they are all equivalent to (9). The expression in (12) and (13) is obtained using a subspace projection idea. This is different from the matrix dilation point of view in [1, 2, 3].*

The equivalence in Proposition 3.2 implies the following statement, connecting the generalized matrix approximation problem in (1) and its classical version:

Corollary 3.1. *Problem (1) is equivalent to*

$$\begin{array}{ll} \underset{\check{X}}{\text{minimize}} & \text{rank}(\check{X}) \\ \text{subject to} & \left\|\check{A}+\check{X}\right\|_2 < 1, \end{array} \tag{15}$$

where $\check{A}$ is defined in (13). The equivalence means (a) that the minimizers of the two optimization problems are one-to-one correspondent as defined in (14), and (b) the

minimum ranks of the two problems are the same. Also, an optimal solution to (1) can be obtained as

$$X^\star = -V_B S_B{}^{-1} \left(U_B{}^T \Delta_C U_B\right)^{-\frac{1}{2}} \left[\check{A}\right]_{\mathrm{sve}(\check{A})} \left(V_C{}^T \Delta_B V_C\right)^{-\frac{1}{2}} S_C{}^{-1} U_C{}^T, \tag{16}$$

where the matrices U_B, S_B, V_B are defined in (3), U_C, S_C, V_C are defined in (4), Δ_B, Δ_C are defined in (11) and the rank constrained truncation operation $\left[\check{A}\right]_{\mathrm{sve}(\check{A})}$ is defined in (6).

PROOF. See Appendix. ■

Remark 3.3. *To characterize all optimal solutions to (1), it suffices to characterize all optimal solutions to (15). The later task is standard (e.g. [9]).*

3.2. Main Result

While Corollary 3.1 provides an SVD based solution expression for the generalized matrix approximation problem in (1), it does not provide an intuitive relationship between the rank of $X^\star$ and the original problem data A, B and C. This is to be complemented by the main result as follows.

Theorem 3.1. *Let the data matrices be defined in the statement of Proposition 3.1. Consider the following generalized minimum rank matrix approximation problem (i.e., problem (1)):*

$$\begin{array}{ll} \underset{X}{\text{minimize}} & \mathrm{rank}(X) \\ \text{subject to} & \|A + BXC\|_2 < 1. \end{array} \tag{17}$$

If the above problem is feasible (i.e., (10) is true), then the minimum rank of the problem in (17) is $\mathrm{sve}(A)$*, where* $\mathrm{sve}(A)$ *is the singular value excess of A (i.e., the number of singular values of A which are greater than or equal to one, see (8)).*

Remark 3.4. *Theorem 3.1 provides a simple characterization of the minimum rank of (17) in terms of* $\mathrm{sve}(A)$*, and states that B and C affect the optimization problem only through the feasibility condition in (10). No analogous result is known for the case where the spectral norm in (17) is replaced with the Frobenius norm.*

Remark 3.5. *Theorem 3.1 states that, under the feasibility assumption in (10), the minimum rank of* $X \in \mathbb{R}^{m_X \times n_X}$ *is* $\mathrm{sve}(A)$ *with* $A \in \mathbb{R}^{m \times n}$*. Since it is assumed in (1.1) that* $m > m_X$ *and* $n > n_X$*, can a contradiction arise that* $\mathrm{rank}(X) = \mathrm{sve}(A) > \min\{m_X, n_X\}$*? Fortunately the answer is no. In particular, the proof of Theorem 3.1 (cf. (19)) implies, under the assumption in (10), that* $\mathrm{sve}(A) = \mathrm{sve}(\check{A}) \leq \min\{m_X, n_X\}$*. In another words,*

$$\max\left\{\left\|N_B{}^T A\right\|_2, \left\|A N_C\right\|_2\right\} < 1 \quad \Longrightarrow \quad \sigma_{(\min\{m_X,\, n_X\}+1)}(A) < 1. \tag{18}$$

Now the proof of Theorem 3.1 begins.

PROOF. As it was argued in the proof of Corollary 3.1, the optimal rank in (17) is the same as that of its equivalence (15), which is $\mathrm{sve}(\check{A})$. To complete the proof, it remains to show that $\mathrm{sve}(\check{A}) = \mathrm{sve}(A)$. 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 $k_-(I - \check{A}\check{A}^T) = k_-(I - AA^T)$. This proof is divided into two steps via an intermediate matrix $\tilde{A}$ defined in the proof of Proposition 3.2:

$$k_-(I - AA^T) = k_-(I - \tilde{A}^T\tilde{A}) = k_-(I - \check{A}\check{A}^T). \tag{19}$$

With the definition of $\tilde{A}$ in (33), the term $k_-(I - \tilde{A}^T\tilde{A})$ in the first equality in (19) becomes $k_-(I - V_C{}^T A^T \Delta_C A V_C) = k_-(I - A V_C V_C{}^T A^T \Delta_C)$, where the later 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. Using the definition and invertibility of Δ_C in (13), the term further becomes $k_-(((\Delta_C)^{-1} - A V_C V_C{}^T A^T)\Delta_C) = k_-((I - A(N_C N_C{}^T + V_C V_C{}^T)A^T)\Delta_C)$. With the identity $N_C N_C{}^T + V_C V_C{}^T = I$ in (5), the above term simplifies to $k_-((\Delta_C)^{\frac{1}{2}}(I - AA^T)(\Delta_C)^{\frac{1}{2}})$. Finally, 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

$$k_-(I - \tilde{A}^T\tilde{A}) = k_-((\Delta_C)^{\frac{1}{2}}(I - AA^T)(\Delta_C)^{\frac{1}{2}}) = k_-(I - AA^T). \tag{20}$$

This shows the first equality in (19). Next, the second equality in (19) can be proved in similar fashions. In particular, using the following four items: (a) The definition of $\check{A}$ in (13), (b) The definition of $\tilde{A}$ in (33), (c) The expression of $\Delta_{\tilde{B}}$ in (37) in Section 6.1 (proved in Section 6.2) and (d) The expression of $U_{\tilde{B}}$ in (39) in Section 6.1 (proved in

Section 6.3), the matrix $\check{A}$ in (19) can be represented as

$$\check{A} = \left(\left(U_B{}^T \Delta_C U_B\right)^{-\frac{1}{2}} U_B{}^T (\Delta_C)^{\frac{1}{2}} \right) \left((\Delta_C)^{\frac{1}{2}} A V_C \right) \left(\left(V_C{}^T \Delta_B V_C\right)^{\frac{1}{2}} \right) = Q U_{\tilde{B}}{}^T \tilde{A} (\Delta_{\tilde{B}})^{\frac{1}{2}},$$

where Q in the above expression is an orthogonal matrix whose exact value is not relevant. Using the above expression of $\check{A}$, the last term in (19) can be written as $k_-(I - Q U_{\tilde{B}}{}^T \tilde{A} \Delta_{\tilde{B}} \tilde{A}^T U_{\tilde{B}} Q^T) = k_-(I - U_{\tilde{B}}{}^T \tilde{A} \Delta_{\tilde{B}} \tilde{A}^T U_{\tilde{B}})$. By expanding $U_{\tilde{B}}$ and $\tilde{A}$, a similar statement as in the case of (20) shows that

$$k_-(I - \tilde{A}^T \tilde{A}) = k_-((\Delta_{\tilde{B}})^{\frac{1}{2}} (I - \tilde{A}^T \tilde{A}) (\Delta_{\tilde{B}})^{\frac{1}{2}}) = k_-(I - \check{A} \check{A}^T). \tag{21}$$

Combining (20) and (21) leads to (19). This concludes the proof of the main result. ■

4. Constrained Generalized Matrix Approximation Problem: SVD Solution

This section describes an SVD based solution procedure for a constrained version of (1), which will be defined in (26). To arrive at this conclusion, a preliminary result based on the work in [4] should be described first.

4.1. Preliminary: SVD solution for a constrained version of (2)

For any matrices $M \in \mathbb{R}^{p \times q_2}$ and $L \in \mathbb{R}^{p \times q_1}$ such that L has full column rank ($= q_1$), consider the following problem:

$$\begin{array}{ll} \underset{X}{\text{minimize}} & \text{rank}\left(\begin{bmatrix} -L & X \end{bmatrix}\right) \\ \text{subject to} & \|M + X\|_2 < 1. \end{array} \tag{22}$$

This problem is a variant of (2), by replacing $\text{rank}(X)$ in (2) with $\text{rank}\left(\begin{bmatrix} -L & X \end{bmatrix}\right)$. Using the result in [4], the above problem can be solved as follows. Denote

$$\begin{array}{rcll} L & = & USV^T & \text{as the SVD of } L \\ P_L M & = & UU^T M & \text{as } M \text{ projected on the range of } L \\ P_L^{\perp} M & = & M - P_L M & \text{as the orthogonal complement of } P_L M \end{array}$$

Then it is claimed that the achievable minimum rank in (22) is $q_1 + \text{sve}(P_L^{\perp} M)$, and an optimal solution can be constructed as

$$X^{\star} = -\left(P_L M + \left[P_L^{\perp} M \right]_{\text{sve}(P_L^{\perp} M)} \right). \tag{23}$$

To see the assertion, note that by [4], for any $k \geq q_1$ it holds that

$$\underset{X \,\middle|\, \operatorname{rank}\left(\begin{bmatrix} -L & X \end{bmatrix}\right) \leq k}{\operatorname{argmin}} \|M+X\|_2 = -\left(P_L M + \left[P_L^{\perp} M\right]_{k-q_1}\right). \tag{24}$$

Therefore, using (24) and the fact that $M = PM + P^{\perp}M$, it can be verified that

$$\min_{X \,\middle|\, \operatorname{rank}\left(\begin{bmatrix} -L & X \end{bmatrix}\right) \leq k} \|M+X\|_2 = \left\| P_L^{\perp} M - \left[P_L^{\perp} M\right]_{k-q_1} \right\|_2 = \sigma_{k-q_1+1}\left(P_L^{\perp} M\right). \tag{25}$$

For any integer k, it is an upper bound of the achievable minimum rank in (22) if and only if k renders the last term in (25) less than one. Therefore, the minimum upper bound, denoted as $k^{\star}$, satisfies the condition that $k^{\star} - q_1 + 1$ is the index of the largest singular value of $P_L^{\perp} M$ which is less than one. In other words, the minimum rank of (22) is $k^{\star} = q_1 + \operatorname{sve}(P_L^{\perp} M)$. Finally, substituting the expression of $k^{\star}$ into (24) gives rise to the solution in (23).

4.2. Result

The equivalence in Proposition 3.2 (or any of its alternatives) provides an SVD based solution to the following constrained generalized matrix approximation problem.

Theorem 4.1. *Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times m_X}$, $C \in \mathbb{R}^{n_X \times n}$ satisfy assumption 1.1. In addition, let $M_1 \in \mathbb{R}^{m_X \times n_{X1}}$, $X_2 \in \mathbb{R}^{m_X \times n_{X2}}$ and $n_X = n_{X1} + n_{X2}$. Partition C into $C^T = \begin{bmatrix} C_1{}^T & C_2{}^T \end{bmatrix}$, with $C_1 \in \mathbb{R}^{n_{X1} \times n}$ and $C_2 \in \mathbb{R}^{n_{X2} \times n}$.*

Assume the data (A, B, C, M_1) are chosen such that the following optimization problem is feasible:

$$\begin{aligned} \underset{X_2}{\text{minimize}} \quad & \operatorname{rank}\left(\begin{bmatrix} M_1 & X_2 \end{bmatrix}\right) \\ \text{subject to} \quad & \left\| A + B \begin{bmatrix} M_1 & X_2 \end{bmatrix} C \right\|_2 < 1. \end{aligned} \tag{26}$$

Then (26) is equivalent to

$$\begin{aligned} \underset{\check{X}}{\text{minimize}} \quad & \operatorname{rank}\left(\begin{bmatrix} -\check{L} & \check{X} \end{bmatrix}\right) \\ \text{subject to} \quad & \left\| \check{A} + \check{X} \right\|_2 < 1, \end{aligned} \tag{27}$$

where

$$
\begin{aligned}
\check{L} &\triangleq -(P_L)^{-1}M_1 \\
\check{A} &\triangleq \left(U_B^T\Delta_{C_2}U_B\right)^{-\frac{1}{2}}U_B^T\Delta_{C_2}(A+BM_1C_1)V_{C_2}\left(V_{C_2}^T\Delta_B V_{C_2}\right)^{\frac{1}{2}} \\
\check{X} &\triangleq (P_L)^{-1}X_2(P_R)^{-1} \\
P_L &\triangleq V_B {S_B}^{-1}\left({U_B}^T\Delta_{C_2}U_B\right)^{-\frac{1}{2}} \\
P_R &\triangleq \left({V_{C_2}}^T\Delta_B V_{C_2}\right)^{-\frac{1}{2}}{S_{C_2}}^{-1}{U_{C_2}}^T \\
B &= \begin{bmatrix}U_B & N_B\end{bmatrix}\begin{bmatrix}S_B\\0\end{bmatrix}{V_B}^T \quad \textit{is the SVD of } B \\
C_2 &= U_{C_2}\begin{bmatrix}S_{C_2}\\0\end{bmatrix}\begin{bmatrix}V_{C_2} & N_{C_2}\end{bmatrix}^T \quad \textit{is the SVD of } C_2 \\
\Delta_B &\triangleq \left(I_n-(A+BM_1C_1)^T N_B{N_B}^T(A+BM_1C_1)\right)^{-1} \\
\Delta_{C_2} &\triangleq \left(I_m-(A+BM_1C_1)N_{C_2}{N_{C_2}}^T(A+BM_1C_1)^T\right)^{-1}.
\end{aligned}
\tag{28}
$$

PROOF. Optimization problem (26) can be written as

$$
\begin{aligned}
&\underset{X_2}{\text{minimize}} \quad \operatorname{rank}\left(\begin{bmatrix}M_1 & X_2\end{bmatrix}\right) \\
&\text{subject to} \quad \|(A+BM_1C_1)+BX_2C_2\|_2<1.
\end{aligned}
\tag{29}
$$

The constraint in (29) has the same form as the inequality in (9). Under the feasibility assumption, this constraint is equivalent to (12) as specified by Proposition 3.2. Therefore, the problem in (29) is equivalent to

$$
\begin{aligned}
&\underset{\check{X}}{\text{minimize}} \quad \operatorname{rank}\left(\begin{bmatrix}M_1 & P_L\check{X}P_R\end{bmatrix}\right) \\
&\text{subject to} \quad \left\|\check{A}+\check{X}\right\|_2<1,
\end{aligned}
\tag{30}
$$

with $\check{A}$, $\check{X}$, P_L, P_R given in (28). The desired statement is resulted by noting that in (30) P_L and P_R are invertible and left and right multiplying invertible matrices does not change the rank of a matrix. ■

Remark 4.1. *Problem (27) has the same form as (22), and hence the solution expression in (23) can be applied. Once a solution $\check{X}^\star$ is found, the expression from (28) can be used to find an optimal solution to (26) as $X_2{}^\star = P_L\check{X}^\star P_R$.*

5. Conclusion

Under feasibility assumption, the generalized matrix approximation problem in (1) is similar to its classical version in (2). (1) possesses its equivalent "classical" form in (15). In addition, the minimum rank of (1) is sve(A), the singular value excess of A. This is analogous to the minimum rank in the classical case in (2). A more general constrained version of (1), as described in (26), turns out to be SVD solvable as well. Even though no simple minimum rank characterization can be reported in this case. The practical applications of the results in this paper, not discussed here, are described in [10, 11].

Acknowledgements

Support from the Swedish Research Council through the Linnaeus Center LCCC is gratefully acknowledged.

6. Appendix

6.1. Proof of Proposition 3.2

The general idea of the proof is that (9) will be shown, successively, to be equivalent to some intermediate inequalities until (12) is finally reached. To begin, note that because of (3), (4) and (5), inequality (9) is equivalent to

$$(AV_C{V_C}^T + BXC)(AV_C{V_C}^T + BXC)^T \prec I - AN_C{N_C}^T A^T = (\Delta_C)^{-1}, \tag{31}$$

where the last equality is due to (11), and it is valid because of the assumption in (10). Inequality (31) is equivalent to $\left\|(\Delta_C)^{\frac{1}{2}}AV_C + (\Delta_C)^{\frac{1}{2}}BXU_CS_C\right\|_2 < 1$, after some algebraic manipulations. Rewrite the above inequality in terms of new notations

$$\left\|\tilde{A} + \tilde{B}\tilde{X}\right\|_2 < 1 \tag{32}$$

with

$$\tilde{A} \triangleq (\Delta_C)^{\frac{1}{2}}AV_C \quad \text{and} \quad \tilde{B} \triangleq (\Delta_C)^{\frac{1}{2}}B \quad \text{and} \quad \tilde{X} \triangleq XU_CS_C. \tag{33}$$

Before the next step of the proof, certain notations need to be introduced first. Since Δ_C is invertible, $\tilde{B}$ in (33) has the same dimension and rank as B assumed in (1.1). Therefore, the SVD of $\tilde{B}$ can be written as

$$\tilde{B} = \begin{bmatrix} U_{\tilde{B}} & N_{\tilde{B}} \end{bmatrix} \begin{bmatrix} S_{\tilde{B}} \\ 0 \end{bmatrix} {V_{\tilde{B}}}^T = U_{\tilde{B}} S_{\tilde{B}} {V_{\tilde{B}}}^T$$

$$\begin{array}{lll} \text{such that} & U_{\tilde{B}} \in \mathbb{R}^{m \times m_X}, & {U_{\tilde{B}}}^T U_{\tilde{B}} = I_{m_X} \\ & N_{\tilde{B}} \in \mathbb{R}^{m \times (m-m_X)}, & {N_{\tilde{B}}}^T N_{\tilde{B}} = I_{m-m_X} \\ & S_{\tilde{B}} \in \mathbb{R}^{m_X \times m_X}, & \text{diagonal and positive definite} \\ & V_{\tilde{B}} \in \mathbb{R}^{m_X \times m_X}, & {V_{\tilde{B}}}^T V_{\tilde{B}} = I_{m_X}. \end{array} \tag{34}$$

By the definition of SVD, $\begin{bmatrix} N_{\tilde{B}} & U_{\tilde{B}} \end{bmatrix}$ is an orthogonal matrix and hence

$${U_{\tilde{B}}}^T N_{\tilde{B}} = 0 \quad \text{and} \quad N_{\tilde{B}} {N_{\tilde{B}}}^T + U_{\tilde{B}} {U_{\tilde{B}}}^T = \begin{bmatrix} N_{\tilde{B}} & U_{\tilde{B}} \end{bmatrix} \begin{bmatrix} N_{\tilde{B}} & U_{\tilde{B}} \end{bmatrix}^T = I. \tag{35}$$

Now the proof of the equivalence between (9) and (12) can be resumed, with the starting point being (32). From (34) and (35) it can be seen that (32) is equivalent to

$$\left(U_{\tilde{B}} {U_{\tilde{B}}}^T \tilde{A} + \tilde{B}\tilde{X} \right)^T \left(U_{\tilde{B}} {U_{\tilde{B}}}^T \tilde{A} + \tilde{B}\tilde{X} \right) \prec I - \tilde{A}^T N_{\tilde{B}} {N_{\tilde{B}}}^T \tilde{A}. \tag{36}$$

It can be shown (in Section 6.2) that, under the assumption in (10), the term $I - \tilde{A}^T N_{\tilde{B}} {N_{\tilde{B}}}^T \tilde{A}$ in the right-hand-side of (36) is positive-definite, and its inverse, denoted as $\Delta_{\tilde{B}}$ can be described by the "non-tilde" matrices as

$$\Delta_{\tilde{B}} \triangleq \left(I - \tilde{A}^T N_{\tilde{B}} {N_{\tilde{B}}}^T \tilde{A} \right)^{-1} = {V_C}^T \Delta_B V_C \succ 0. \tag{37}$$

Then, multiplying both sides of (36) with $(\Delta_{\tilde{B}})^{\frac{1}{2}}$, expanding $\tilde{B}$ as $\tilde{B} = U_{\tilde{B}} S_{\tilde{B}} {V_{\tilde{B}}}^T$, and simplifying using the relationship ${U_{\tilde{B}}}^T U_{\tilde{B}} = I$, inequality (36) becomes

$$\left\| {U_{\tilde{B}}}^T \tilde{A} (\Delta_{\tilde{B}})^{\frac{1}{2}} + S_{\tilde{B}} {V_{\tilde{B}}}^T \tilde{X} (\Delta_{\tilde{B}})^{\frac{1}{2}} \right\|_2 < 1. \tag{38}$$

To obtain (12) with $\check{A}$ and $\check{X}$ represented by the original "non-tilde" matrices as in (13). The following expressions (proved in Section 6.3) are needed.

$$\begin{array}{rcl} U_{\tilde{B}} & = & (\Delta_C)^{\frac{1}{2}} U_B \left({U_B}^T \Delta_C U_B \right)^{-\frac{1}{2}} Q \\ S_{\tilde{B}} {V_{\tilde{B}}}^T & = & Q^T \left({U_B}^T \Delta_C U_B \right)^{\frac{1}{2}} S_B {V_B}^T \\ N_{\tilde{B}} & = & (\Delta_C)^{-\frac{1}{2}} N_B \left({N_B}^T (\Delta_C)^{-1} N_B \right)^{-\frac{1}{2}} Q_1, \end{array} \tag{39}$$

where Q and Q_1 are orthogonal matrices whose exact forms are irrelevant to the discussion in here. Using the expressions of the "tilde" quantities in (39), (33) and (37), inequality (38) becomes

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

with Q being a unspecified orthogonal matrix. However, since the spectral norm is unitarily invariant, the above inequality is equivalent to the one without Q. This is the same as (12) with $\check{A}$ defined in (13) and $\check{X}$ defined in (14). Finally, the one-to-one correspondence and its inverse in (14) can be obtained from the above expression as both $\left(U_B{}^T \Delta_C U_B\right)^{\frac{1}{2}} S_B V_B{}^T$ and $U_C S_C \left(V_C{}^T \Delta_B V_C\right)^{\frac{1}{2}}$ are invertible. ■

6.2. Proof of the expression in (37)

Using the definition of Δ_B in (11), the matrix $V_C{}^T \Delta_B V_C \succ 0$ is expanded into

$$V_C{}^T \Delta_B V_C = V_C{}^T \left(I - A^T N_B N_B{}^T A\right)^{-1} V_C = V_C{}^T \left(I - A^T N_B \left(N_B{}^T A A^T N_B - I\right)^{-1} N_B{}^T A\right) V_C,$$

with the second equality due to the matrix inversion lemma [12]. Using the definition of Δ_C in (11) and the identity $V_C V_C{}^T + N_C N_C{}^T = I$ in (5), the last term becomes

$$I - V_C{}^T A^T N_B \left(N_B{}^T (A V_C V_C{}^T A^T - (\Delta_C)^{-1}) N_B\right)^{-1} N_B{}^T A V_C.$$

With another application of the matrix inversion lemma, the above term becomes

$$\left(I - V_C{}^T A^T N_B \left(N_B{}^T (\Delta_C)^{-1} N_B\right)^{-1} N_B{}^T A V_C\right)^{-1} = \left(I - \tilde{A}^T N_{\tilde{B}} N_{\tilde{B}}{}^T \tilde{A}\right)^{-1},$$

where the last equality is due to the definition of $\tilde{A}$ in (33) and the expression of $N_{\tilde{B}}$ in (39), which will be shown next. ■

6.3. Proof of the expressions in (39)

To show the first line of (39), notice from (34), (33), (3) that the SVD of $\tilde{B}$ is

$$\tilde{B} = U_{\tilde{B}} S_{\tilde{B}} V_{\tilde{B}}{}^T = (\Delta_C)^{\frac{1}{2}} U_B S_B V_B{}^T = (\Delta_C)^{\frac{1}{2}} B. \tag{40}$$

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

$$U_{\tilde{B}} = (\Delta_C)^{\frac{1}{2}} U_B P, \tag{41}$$

with P being an invertible matrix. By the definition of $U_{\tilde{B}}$ in (34), 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 an orthogonal matrix Q such that $P = (U_B{}^T \Delta_C U_B)^{-\frac{1}{2}} Q$. Substituting the above expression into (41), $U_{\tilde{B}}$ yields the first line in (39).

From the second equality in (40) and the expression of $U_{\tilde{B}}$ in the first line in (39) it can be seen that

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

This is the same as the second line in (39).

To show the third line of (39), the relations in (35) and the first line in (39) imply 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 $U_B{}^T (\Delta_C)^{\frac{1}{2}} N_{\tilde{B}} = 0$. Hence, $(\Delta_C)^{\frac{1}{2}} N_{\tilde{B}}$ is in the kernel of $U_B{}^T$, and there exists a square matrix Y such that

$$N_{\tilde{B}} = (\Delta_C)^{-\frac{1}{2}} N_B Y. \tag{42}$$

Also, by the definition of $N_{\tilde{B}}$ in (34), $N_{\tilde{B}}{}^T N_{\tilde{B}} = Y^T N_B{}^T (\Delta_C)^{-1} N_B Y = I$. Since the matrix $(N_B{}^T (\Delta_C)^{-1} N_B)^{\frac{1}{2}}$ is square, the above identity implies that there exists an orthogonal 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 (42) yields the third line in (39). ■

6.4. Proof of Corollary 3.1

The equivalence between the optimization problems in (1) and (15) is a consequence of the equivalence of the inequalities in (9) and (12), as well as the correspondence in (14). Since an optimal solution to the classical problem (15) is $-\left[\check{A}\right]_{\mathrm{sve}(\check{A})}$, an application of (14) results in the desired expression in (16). ■

References

[1] K. Zhou, J. Doyle, K. Glover, Robust and Optimal Control, Prentice Hall, 1996.

[2] C. Davis, W.M. Kahan, H.F. Weinberger, Norm-preserving dilations and their applications to optimal error bounds, SIAM Journal on Numerical Analysis 19 (3) (1982) pp. 445–469.

[3] A. Megretski, MIT 6.245 Multivariable Control Systems, website: http://web.mit.edu/6.245/www/.

[4] G. Golub, A. Hoffman, G. Stewart, A Generalization of the Eckart-Young-Mirsky Matrix Approximation Theorem, Linear Algebra and its Applications 88/89 (1987) 317–327.

[5] J. Demmel, The smallest perturbation of a submatrix which lowers the rank and constrained total least squares problems, SIAM Journal on Numerical Analysis 24 (1) (1987) 206.

[6] H. Zha, A numerical algorithm for computing the restricted singular value decomposition of matrix tripiets, Linear Algebra and its Applications 168 (1992) 1–25.

[7] Y. Nievergelt, Schmidt-Mirsky Matrix Approximation With Linearly Constrained Singular Values, Linear Algebra and its Applications 261 (1997) 207–219.

[8] S. Friedland, A. Torokhti, Generalized rank-constrained matrix approximations, SIAM Journal on Matrix Analysis and Applications 29 (2) (2006) 656–659.

[9] R. Horn, C. Johnson, Matrix Analysis, Cambridge University Press, 1990.

[10] K.C. Sou, A. Rantzer, A singular value decomposition based closed loop stability preserving controller reduction method, in: American Control Conference, 2010.

[11] K.C. Sou, A. Rantzer, Controller reduction with closed loop error guarantee, in: Conference on Decision and Control, 2010.

[12] N. Higham, Accuracy and Stability of Numerical Algorithms, SIAM, 2002.