R.A. Delgado, J.C. Aüero and G.C. Goodwin. **A novel representation of rank constraints for real matrices**. *Linear Algebra and its Applications*, Vol 496:452-462, May 2016.

Available at http://dx.doi.org/10.1016/j.laa.2016.02.008

A novel representation of rank constraints for real matrices☆

Ramón A. Delgado[a,*], Juan C. Agüero[b], Graham C. Goodwin[a]

[a] *School of Electrical Engineering and Computer Science, The University of Newcastle, Australia*
[b] *Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile*

Abstract

We present a novel representation of rank constraints for non-square real matrices. We establish relationships with existing results and show that these are particular cases of our representation. One of these cases, is a representation of the ℓ_0 pseudo-norm, which is used in sparse representation problems. Finally, we describe how our representation can be included in rank-constrained optimization and in rank-minimization problems.

Keywords: rank-minimisation, low-rank approximation, sparse representation.
2010 MSC: 15A03, 15A23, 65F30, 90C27, 93A30.

1. Introduction

Rank constraints find application in many areas including data modelling, systems and control, computer algebra, signal processing, psychometrics, machine learning, computer vision, among others [1, 2]. In many applications, the notion of complexity of a model can be related to the rank of a particular matrix. For example, in factor analysis, the number of latent factors is equal to

☆Juan C. Agüero was partially supported by the Chilean Research Council (CONICYT) through Basal Project FB0008, and FONDECYT (grant no. 1150954).

*Corresponding author

Email addresses: `ramon.delgado@uon.edu.au` (Ramón A. Delgado), `juan.aguero@usm.cl` (Juan C. Agüero), `graham.goodwin@newcastle.edu.au` (Graham C. Goodwin)

the rank of a covariance matrix. In system identification, the order of a rational system is equal to the rank of an infinite dimensional Hankel matrix.

Handling rank constraints is known to be difficult since the rank$(\cdot)$ function has undesirable features. In particular, the function is non-smooth, non-linear and non-convex. In applications based on optimization, where smoothness and convexity are widely exploited, the non-smoothness of the rank$(\cdot)$ function limits the tools that can be used in the optimization problem. On the other hand, non-smoothness is often tolerated in order to obtain other desirables properties, e.g. convexity, in some other part of the problem. This has motivated several authors [3, 4, 5, 1, 6] to find equivalent representations for rank constraints. These representations are equivalent ways to express a rank constraint of the form $\text{rank}(A) \leq r$. These equivalent representations are aimed at overcoming the non-linearity, non-smoothness and/or non-convexity of the rank function. For example, one equivalent representation of a rank constraint for $A \in \mathbb{R}^{m\times n}$ is given by

$$\text{rank}(A) \leq r \iff \exists \text{ a full row rank matrix } U \in \mathbb{R}^{(m-r)\times m} \text{ such that } UA = 0 \tag{1}$$

One advantage of the rank representation (1) is that it frees the matrix A to satisfy other structural constraint. In [1] the rank representation (1) has been used to impose a rank constraint in a structured matrix, such as a Hankel matrix. However, this rank constraint representation have some limitations. First, it transfers a rank constraint from a matrix, A, to an auxiliary matrix U. Moreover, the rank of the matrix A is related to the size of the auxiliary matrix U. This last issue makes this approach inappropriate for problems where the rank to be constrained is selected in a dynamic way, e.g. online.

Other existing rank constraint representations are valid only for positive semidefinite matrices, see e.g. [3, 5, 4]. For example, consider the rank constraint representation in [3] that establishes that, for a matrix $A \in \mathbb{S}^n_+$, then

$$\text{rank}(A) \leq r \iff \exists W \in \Phi_{n,r} \text{ such that } \text{trace}(AW) = 0. \tag{2}$$

where

$$\Phi_{n,r} = \{W \in \mathbb{S}^n,\ 0 \preceq W \preceq I, \text{trace}(W) = n - r\} \tag{3}$$

This rank constraint representation eliminates of the rank function, but is only valid for positive semidefinite matrices. A detailed discussion of these and other rank representations is given in section 2.

In this paper we present a novel representation of rank constraints. This novel representation is valid for all real matrices and can be seen as a generalisation of existing methods. Moreover, the representation overcomes many of the limitations associated with existing approaches.

The remainder of the paper is organised as follows: The main result is presented in Section 2. Section 3 establishes connections with existing results. Section 4 shows how the result can be applied to a class of optimization problems. Finally, conclusions are drawn in Section 5.

Notation and basic definitions: $\text{rank}(A)$ denotes the rank of a matrix A. We denote by $A^\dagger$ the Moore-Penrose pseudoinverse of A. $\lambda_i(A)$ denotes the i-th largest eigenvalue of a symmetric matrix A and $\sigma_i(A)$ denotes the i-th largest singular value of a matrix A. $A \succeq 0$ denotes that A is positive semidefinite, and $A \succeq B$ denotes that $A - B \succeq 0$. We denote the transpose of a given matrix A as $A^\top$. $\mathbb{S}^n$ denotes the set of symmetric matrices of size $n \times n$, and $\mathbb{S}^n_+$ the set of symmetric positive semidefinite matrices, i.e. $\mathbb{S}^n_+ := \{A \in \mathbb{S}^n | A \succeq 0\}$. $\|\cdot\|_F = \sqrt{\text{trace}(A^\top A)}$ denotes the Frobenius norm and $\|\cdot\|_* = \text{trace}(\sqrt{A^\top A})$ denotes the nuclear norm, a.k.a. trace norm. $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm that counts the number of nonzero elements of a vector.

2. Main Result

The main result of this paper is the following:

Theorem 1. *Let $A \in \mathbb{R}^{m\times n}$, then the following expressions are equivalent*

(i) $\text{rank}(A) \leq r$

(ii) $\exists W_R \in \Phi_{n,r}$, *such that* $AW_R = 0_{m\times n}$

(iii) $\exists W_L \in \Phi_{m,r}$, *such that* $W_L A = 0_{m\times n}$

where

$$\Phi_{n,r} = \{W \in \mathbb{S}^n,\ 0 \preceq W \preceq I, \mathrm{trace}(W) = n - r\} \tag{4}$$

Proof:. See Appendix B. □

To the best of the authors' knowledge, Theorem 1 is novel. The closest result on rank-constrained optimization is described by equation (1). The earlier constraint (1) requires of an auxiliary matrix U having full row rank. In optimization problems, this requirement may lead to the necessity of including additional non-convex constraints, such as $UU^\top = I_{m-r}$.

Another closely related result was introduced in [3, §4.4] and is given by the rank representation (2)-(3). The latter result makes use of the convex set $\Phi_{n,r}$ in (3), but the formalism is valid only for positive semidefinite matrices. Notice that Theorem 1 can be seen as a generalisation of (2) to non-positive semidefinite matrices.

There exist other rank-constraint representations which impose conditions on the coefficients of the characteristic polynomial of the matrix, see e.g. [4, 5]. These are also valid only for positive semidefinite matrices. Additionally, these rank constraints representations have the disadvantage that, in general, computing the coefficients of the characteristic polynomial of a matrix is not an easy task.

Notice that one of the key steps in proving Theorem 1 is the observation that for all $W \in \mathbb{S}^n$ such that $0 \preceq W \preceq I$, it is true that $\mathrm{trace}(W) \leq \mathrm{rank}(W)$. This fact is a consequence of a stronger result that states that, in the set of interest, $\{W \in \mathbb{S}^n | 0 \preceq W \preceq I\}$, the trace function is the largest convex function that is less than or equal to the rank function. This latter result is one of the key underlying ingredients in the development of the nuclear norm heuristic [7].

3. Connection to existing results

In this section we establish connections between Theorem 1 and other existing results. The following lemma establishes the relationship between Theorem 1 and the rank-constraint representation (2).

Lemma 1. *Let* $A \in \mathbb{S}^n_+$ *and* $W \in \mathbb{S}^n_+$*, then*

$$\operatorname{trace}(AW) = 0 \iff AW = 0 \tag{5}$$

Proof:. Since the matrices A and W are both symmetric and positive semidefinite, then by the *Cholesky decomposition*, see e.g. [8, Fact 8.9.37], there exist matrices $P \in \mathbb{R}^{n\times n}$ and $Q \in \mathbb{R}^{n\times n}$ such that

$$A = PP^\top \tag{6}$$

$$W = QQ^\top \tag{7}$$

We then have that

$$\begin{aligned} \operatorname{trace}(AW) &= \operatorname{trace}(PP^\top QQ^\top) & (8)\\ &= \operatorname{trace}(Q^\top PP^\top Q) & (9) \end{aligned}$$

Next, we recall that, for $B \in \mathbb{R}^{m\times n}$, the *Frobenius norm* is defined by $\|B\|_F = \sqrt{\operatorname{trace}(B^\top B)}$, see e.g. [8, page 601]. Then, we have

$$\operatorname{trace}(Q^\top PP^\top Q) = \|P^\top Q\|_F^2 \tag{10}$$

and from the definition of a norm we have that $\|B\| = 0$ if and only if $B = 0$, see e.g. [8, Definition 9.2.1.]. Then we have

$$\operatorname{trace}(AW) = \|P^\top Q\|_F^2 = 0 \iff AW = 0 \tag{11}$$

This concludes the proof. □

A recent result in [9] is closely related to the rank representation in Theorem 1. However, in [9] the result for general real matrices is obtained via matrix augmentation. This approach introduce several auxiliary matrices into the rank-constraint representation.

Another particular case of Theorem 1 is a representation of the ℓ_0 pseudo norm, denoted as $\|\cdot\|_0$. This corresponds to the number of non-zero elements. The connection is made by considering a diagonal matrix $A \in \mathbb{R}^{n\times n}$ such that its diagonal elements are given by a vector $x \in \mathbb{R}^n$, i.e. $A = \text{diag}\{x\}$. We have that $\|x\|_0 = \text{rank}(A)$. Then, Theorem 1 can be used to prove the following result.

Corollary 1. *Let $x \in \mathbb{R}^n$, then the following expressions are equivalent*

(i) $\|x\|_0 \leq r$

(ii) $\exists w \in \{w \in \mathbb{R}^n |\ 0 \leq w_i \leq 1, i = 1, \ldots, n; \sum_{i=1}^n w_i = n - r\}$, *such that* $x_i w_i = 0$ *for* $i = 1, \ldots, n$.

Proof:. Consider the following definition $A = \text{diag}\{x\} \in \mathbb{R}^{n\times n}$, i.e.

$$A = \begin{bmatrix} x_1 & 0 & 0 & \cdots & 0 \\ 0 & x_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & x_{n-1} & 0 \\ 0 & \cdots & 0 & 0 & x_n \end{bmatrix} \tag{12}$$

Notice that by construction, $\text{rank}(A) = \|x\|_0$. From Theorem 1 we have $\text{rank}(A) \leq r$, if and only if, there exists a $W \in \{W \in \mathbb{S}^n |\ 0 \preceq W \preceq I; \text{trace}(W) = n - r\}$ such that $AW = 0$. Since A is diagonal, then without loss of generality, we can assume that $W = \text{diag}\{w\}$. This can be easily seen by defining $C = AW$ and considering that W is symmetric. Note that, since A is diagonal, $C_{ij} = A_{ii}W_{ij}$ and $C_{ji} = A_{jj}W_{ji}$. If $A_{ii} = A_{jj} = 0$ for $i \neq j$ then $W_{ij} = W_{ji}$ can take any value, including zero, and still satisfy $C_{ij} = C_{ji} = 0$. If $A_{ii} \neq 0$ then $W_{ij} = W_{ji} = 0$ in order to satisfy $C_{ij} = 0$. Finally, the conditions on w are directly derived from conditions on W. $\square$

We note, in passing, that this representation of the ℓ_0 constraint is related to the results reported in [10, 11, 4, 12],[3, §4.5].

4. Applications in Optimization

In this section we show how Theorem 1 can be utilized to include rank constraints into optimization problems. In the last decade there has been increasing interest in the problem of including rank matrix functions in optimization problems. This is motivated by the introduction of the nuclear norm heuristic [7], which provides a convex relaxation for rank-minimization problems. The nuclear norm heuristic has been shown to be particularly useful for high-dimensional optimization problems. However, as has been shown in [13], there is an inherent loss of performance when the nuclear norm heuristic is utilised.

Theorem 1 can be applied to rank-constrained optimization problems by simply replacing the rank contraint by one of the equivalent representations, as follows

$$\begin{aligned} \mathcal{P}_{rco}: \quad \min_{\theta \in \mathbb{R}^p} \quad & f(\theta) \\ \text{s.t.} \quad & \theta \in \Theta \\ & \operatorname{rank}(A(\theta)) \leq r \end{aligned} \quad \equiv \quad \begin{aligned} \mathcal{P}_{rcoequiv}: \quad \min_{\theta \in \mathbb{R}^p, W \in \mathbb{S}^n} \quad & f(\theta) \\ \text{s.t.} \quad & \theta \in \Theta \\ & A(\theta)W = 0_{m \times n} \\ & W \in \Phi_{n,r} \end{aligned}$$

Similarly, Theorem 1 can also be applied to rank-minimization problems by using the epigraph representation [14], as follows

$$\begin{aligned} \mathcal{P}_{rm}: \quad \min_{\theta \in \mathbb{R}^p} \quad & r \\ \text{s.t.} \quad & \theta \in \Theta \\ & \operatorname{rank}(A(\theta)) \leq r \end{aligned} \quad \equiv \quad \begin{aligned} \mathcal{P}_{rmequiv}: \quad \min_{\theta \in \mathbb{R}^p, W \in \mathbb{S}^n} \quad & n - \operatorname{trace}(W) \\ \text{s.t.} \quad & \theta \in \Theta \\ & A(\theta)W = 0_{m \times n} \\ & 0 \preceq W \preceq I_n \end{aligned}$$

The equivalence of the rank-minimization problem $\mathcal{P}_{rm}$ to the problem $\mathcal{P}_{rmequiv}$ can be seen as a generalisation, to general real matrices, of the results presented in [4].

The equivalence between $\mathcal{P}_{rm}$ and $\mathcal{P}_{rmequiv}$ has been applied, by the current authors, to several rank-constrained optimization problems. In order to illustrate the advantages of using our rank-contraint representation, we briefly described the results in [15] where the problem of Factor Analysis has been considered.

In Factor Analysis one is given a set of N measurements $y_t \in \mathbb{R}^n$. It is assumed that the measurements can be decomposed into a common and an idiosyncratic part. The common part is assumed to be a linear combination of a small number of factors, $f_t \in \mathbb{R}^r$ which are normally distributed $f_t \sim N(0, I_r)$, with $r < n$. The idiosyncratic part is assumed to satisfy $v_t \sim N(0, \Psi)$, with $\Psi \in \mathbb{S}^n_+$. Thus, we have that the data-generating process is described by $y_t = Cf_t + v_t$ where $C \in \mathbb{R}^{n \times r}$ denotes the factor loadings. The aim is to obtain an estimate $\hat{C}$ of the factor loadings, and an estimate $\hat{f}_t$ of the factors. The problem can also be expressed as that of decomposing the covariance matrix $\Sigma = \frac{1}{N}\sum_{t=1}^{N} y_t y_t^T$ as follows $\Sigma = C^\top C + \Psi$. Existing methods for Factor Analysis require that Ψ be diagonal. In [15] the rank constraint representation in Theorem 1 has been applied to relax the restriction of diagonal Ψ. Instead, it was assumed in [15] that Ψ is sparse. This relaxation on the assumptions is possible because the rank constraint representation in Theorem 1 allows more freedom on Ψ. The approach in [15] formulates the Factor Analysis problem as the following rank-constrained optimisation problem.

$$
\begin{aligned}
\mathcal{P}_{FA}: \qquad & \min_{\Psi \in \mathbb{S}^n} \ \|\Psi\|_1 \\
& \text{subject to} \ \ \operatorname{rank}(\Sigma - \Psi) \leq r \\
& \qquad\qquad\quad 0 \preceq \Psi \preceq \Sigma
\end{aligned}
$$

In [15] it is shown that, in certain scenarios, the approach $\mathcal{P}_{FA}$ is competitive with state-of-the-art methods for Factor Analysis even when local optimization is used to solve $\mathcal{P}_{FA}$.

The current authors also have applied Corollary 1 to several other areas. For example, Corollary 1 has been used in [16] to impose a ℓ_0 contraint on a Model Predictive Control problem. This Corollary has also been used in [17] to estimate the parameters of a nonlinear dynamic system.

5. Conclusions

We have described a novel representation of rank constraints for real matrices. We have established a link between the representation and several existing results. These have been shown to be particular cases of our representation. One of these particular cases, is a representation for the ℓ_0 pseudo-norm, which is often used in sparse representation problems. We have also described how our representation can be included in rank-constrained optimization and in rank-minimization problems.

Appendix A. Background results

In this section we review several existing results which are deployed in the sequel. These known results are included here for the reader's convenience.

We first present *Sylvester's inequality*, for real matrices. This provides a lower bound for the rank of the product of two matrices.

Proposition 1. *[8, Proposition 2.5.9] Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times l}$. Then,*

$$\operatorname{rank}(A) + \operatorname{rank}(B) \leq m + \operatorname{rank}(AB). \tag{A.1}$$

Next, we review several results pertaining to Singular Value Decompositions.

Definition 1. *[8, Definition 5.6.1.] Let $A \in \mathbb{R}^{m\times n}$. Then the* singular values *of A are the $\min\{m, n\}$ nonnegative numbers $\sigma_1(A), \ldots, \sigma_{\min\{m,n\}}(A)$, where, for all $i = 1, \ldots, \min\{m, n\}$,*

$$\sigma_i(A) \coloneqq \lambda_i^{1/2}(AA^\top) = \lambda_i^{1/2}(A^\top A). \tag{A.2}$$

Hence,

$$\sigma_1(A) \geq \cdots \geq \sigma_{\min\{m,n\}}(A) \geq 0. \tag{A.3}$$

Note that if $A \in \mathbb{S}^n$ and is positive semidefinite, then $\sigma_i(A) = \lambda_i(A)$ for $i = 1, \ldots, n$.

Definition 2. *[8, Definition 3.1.1. xxiv] Let $A \in \mathbb{R}^{n\times n}$, then A is said to be orthonormal if $A^\top A = I_n$*

Theorem 2. *[8, Theorem 5.6.3.] Let $A \in \mathbb{R}^{m\times n}$, assume that A is nonzero. Let $c \coloneqq \operatorname{rank}(A)$, and define $B \coloneqq \operatorname{diag}\{\sigma_1(A), \ldots, \sigma_c(A)\}$. Then, there exists orthonormal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ such that*

$$A = U \begin{bmatrix} B & 0_{c\times(n-c)} \\ 0_{(m-c)\times c} & 0_{(m-c)\times(n-c)} \end{bmatrix} V^\top. \tag{A.4}$$

Appendix B. Proof of Theorem 1

Appendix B.1. Preliminary results

We first present several preliminary results that describe properties of the convex set $\{W \in \mathbb{S}^n | 0 \preceq W \preceq I_n\}$. These properties will be used in the proof of the main result to follow.

Let $A \in \mathbb{R}^{m\times n}$. Assume that A is nonzero. Let $c \coloneqq \operatorname{rank}(A)$, and define $B \coloneqq \operatorname{diag}\{\sigma_1(A), \ldots, \sigma_c(A)\}$. Then, the Singular Value Decomposition (SVD) of A is given by

$$A = U \begin{bmatrix} B & 0_{c\times(n-c)} \\ 0_{(m-c)\times c} & 0_{(m-c)\times(n-c)} \end{bmatrix} V^\top. \tag{B.1}$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthonormal matrices.

Let $r \in \mathbb{N}$ be such that $c \leq r \leq \min\{m, n\}$ and consider the following block partition

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \left[\begin{array}{cc|c} B & 0_{c\times(c-r)} & 0_{c\times(n-r)} \\ 0_{(r-c)\times(c-r)} & 0_{(r-c)\times(c-r)} & 0_{(r-c)\times(n-r)} \\ \hline 0_{(m-r)\times(c-r)} & 0_{(m-r)\times(c-r)} & 0_{(m-r)\times(n-r)} \end{array}\right] \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix}. \tag{B.2}$$

where $U_1 \in \mathbb{R}^{m\times r}$, $U_2 \in \mathbb{R}^{m\times(m-r)}$, $V_1 \in \mathbb{R}^{n\times r}$, and $V_2 \in \mathbb{R}^{n\times(n-r)}$.

Based on the block partition (B.2) of the SVD, we establish the following two Lemmas.

Lemma 2. *Let* $A \in \mathbb{R}^{m\times n}$ *with SVD given by* (B.2), *and let* $W_R \coloneqq V_2 V_2^\top$. *Then,* $W_R \in \mathbb{R}^{n\times n}$ *satisfies*

(i) $W_R \succeq 0$.

(ii) $W_R = W_R^\top$.

(iii) $W_R \preceq I_n$.

(iv) $\mathrm{trace}(W_R) = n - r$.

(v) $AW_R = 0_{m\times n}$.

Proof:. By construction, **(i)** is a well known result, see e.g. [8, Fact 3.7.25]. The proof of **(ii)** is straight forward since $V_2V_2^\top = (V_2V_2^\top)^\top$. To prove **(iii)** , consider in (B.2) that V is an orthonormal matrix, then $V_1^\top V_1 = I_r$, $V_2^\top V_2 = I_{n-r}$, $V_1 V_2^\top = 0_{n\times n}$ and

$$V_1V_1^\top + V_2V_2^\top = I_n \tag{B.3}$$

where $V_1V_1^\top$ is positive semidefinite. Then $I_n - W_R = V_1V_1^\top \succeq 0$. Next we prove **(iv)**. By using the trace operator in (B.3), we have that

$$\begin{aligned} \mathrm{trace}(W_R) &= \mathrm{trace}(I_n) - \mathrm{trace}(V_1V_1^\top) && \text{(B.4)} \\ &= n - \mathrm{trace}(V_1^\top V_1) && \text{(B.5)} \\ &= n - \mathrm{trace}(I_r) = n - r. && \text{(B.6)} \end{aligned}$$

Finally, to prove **(v)**, we note that, by using (B.2), we have

$$A\,W_R = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \left[\begin{array}{cc|c} B & 0_{c\times(c-r)} & 0_{c\times(n-r)} \\ 0_{(r-c)\times(c-r)} & 0_{(r-c)\times(c-r)} & 0_{(r-c)\times(n-r)} \\ \hline 0_{(m-r)\times(c-r)} & 0_{(m-r)\times(c-r)} & 0_{(m-r)\times(n-r)} \end{array}\right] \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix} V_2 V_2^\top \tag{B.7}$$

$$= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \left[\begin{array}{cc|c} B & 0_{c\times(c-r)} & 0_{c\times(n-r)} \\ 0_{(r-c)\times(c-r)} & 0_{(r-c)\times(c-r)} & 0_{(r-c)\times(n-r)} \\ \hline 0_{(m-r)\times(c-r)} & 0_{(m-r)\times(c-r)} & 0_{(m-r)\times(n-r)} \end{array}\right] \begin{bmatrix} V_1^\top V_2 \\ V_2^\top V_2 \end{bmatrix} V_2^\top \tag{B.8}$$

$$= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \left[\begin{array}{cc|c} B & 0_{c\times(c-r)} & 0_{c\times(n-r)} \\ 0_{(r-c)\times(c-r)} & 0_{(r-c)\times(c-r)} & 0_{(r-c)\times(n-r)} \\ \hline 0_{(m-r)\times(c-r)} & 0_{(m-r)\times(c-r)} & 0_{(m-r)\times(n-r)} \end{array}\right] \begin{bmatrix} 0_{r\times(n-r)} \\ I_{n-r} \end{bmatrix} V_2^\top \tag{B.9}$$

$$= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} 0_{r\times(n-r)} \\ 0_{(m-r)\times(n-r)} \end{bmatrix} V_2^\top \tag{B.10}$$

$$= 0_{m\times n} \tag{B.11}$$

$\square$

Lemma 3. *Let $A \in \mathbb{R}^{m\times n}$ with SVD given by* (B.2)*, and let $W_L \coloneqq U_2 U_2^\top$. Then, $W_L \in \mathbb{R}^{m\times m}$ satisfies*

(i) $W_L \succeq 0$

(ii) $W_L \preceq I_m$

(iii) $W_L = W_L^\top$

(iv) $\mathrm{trace}(W_L) = m - r$

(v) $W_L A = 0_{m\times n}$

Proof:. The proof is similar to the proof of Lemma 2. $\square$

Lemma 4. *Let $W \in \mathbb{S}^n$ be such that $0 \preceq W \preceq I_n$, then*

$$\operatorname{trace}(W) \leq \operatorname{rank}(W) \tag{B.12}$$

Proof:. Let $c \coloneqq \operatorname{rank}(W)$, and define $B \coloneqq \operatorname{diag}\{\sigma_1(W), \dots, \sigma_c(W)\}$ and consider the SVD

$$W = U \begin{bmatrix} B & 0_{c\times(n-c)} \\ 0_{(m-c)\times c} & 0_{(m-c)\times(n-c)} \end{bmatrix} U^\top. \tag{B.13}$$

then

$$\operatorname{trace}(W) = \sum_{k=1}^{c} \sigma_i(W) \tag{B.14}$$

Since $W \in \mathbb{S}^n$ and $W \succeq 0$, we have that $\sigma_i(W) = \lambda_i(W)$ for all $i = 1, \dots, n$. Furthermore, since $W \preceq I_n$ we have that $1 \geq \lambda_1(W)$, see [8, Lemma 8.4.1 iii], then $1 \geq \lambda_1(W) \geq \cdots \geq \lambda_c(W) \geq \cdots \geq \lambda_n(W)$. Finally, we have that

$$\operatorname{trace}(W) = \sum_{k=1}^{c} \lambda_i(W) \leq c = \operatorname{rank}(W) \tag{B.15}$$

$\square$

Appendix B.2. Proof of Theorem 1

Proof:. Lemma 2 establishes that **(i)** $\implies$ **(ii)**. Lemma 3 proves that **(i)** $\implies$ **(iii)**

Next, we establish **(ii)** $\implies$ **(i)**. From Lemma 4 we have that

$$\operatorname{trace}(W_R) \leq \operatorname{rank}(W_R) \tag{B.16}$$

On the other hand, by using *Sylvester's Inequality*, we have that

$$\operatorname{rank}(A) + \operatorname{rank}(W_R) \leq n + \operatorname{rank}(AW_R) \tag{B.17}$$

Then, by using (B.16), we have

$$\operatorname{rank}(A) + \operatorname{trace}(W_R) \leq n + \operatorname{rank}(AW_R) \tag{B.18}$$

Then, by using the fact that $\text{rank}(AW_R) = \text{rank}(0_{m\times n}) = 0$, we obtain

$$\text{rank}(A) \leq n - \text{trace}(W_R) \tag{B.19}$$

Since $W_R \in \Phi_{n,r}$, we have that $\text{trace}(W_R) = n - r$. Then

$$\text{rank}(A) \leq r. \tag{B.20}$$

This completes the proof that **(ii)** $\Longrightarrow$ **(i)**. The proof of **(iii)** $\Longrightarrow$ **(i)** is similar to the proof of **(ii)** $\Longrightarrow$ **(i)**. □

References

[1] I. Markovsky, Low Rank Approximation: Algorithms, Implementation, Applications, Communications and Control Engineering, Springer, 2012.

[2] J. Lee, B. Zou, Optimal rank-sparsity decomposition, Journal of Global Optimization 60 (2) (2014) 307–315.

[3] J. Dattorro, Convex optimization and Euclidean distance geometry, Meboo Publishing, USA, 2005.

[4] A. d'Aspremont, A semidefinite representation for some minimum cardinality problems, in: 42nd IEEE Conference on Decision and Control, 2003, Vol. 5, 2003, pp. 4985–4990.

[5] A. Helmersson, On polynomial coefficients and rank constraints, Tech. rep., Linköping University (2009).

[6] V. S. Mai, D. Maity, B. Ramasubramanian, Convex methods for rank-constrained optimization problems, in: Proceedings of the SIAM Conference on Control and its Apllications, Paris, 2015.

[7] M. Fazel, A rank minimization heuristic with application to minimum order system approximation, in: Proceedings of the American Control Conference, IEEE, 2001, pp. 4734–4739.

[8] D. S. Bernstein, Matrix mathematics: theory, facts, and formulas, Princeton University Press, 2009.

[9] A. Hempel, P. Goulart, A novel method for modelling cardinality and rank constraints, in: Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on, 2014, pp. 4322–4327.

[10] J. Mitchell, J.-S. Pang, A. Waechter, L. Bai, M. Feng, X. Shen, Complementarity formulations for ℓ_0 norm optimization problems, in: Presentation at the 11-th Workshop on Advances in Continuous Optimization, Florence, 2013.

[11] D. Piga, R. Tóth, An SDP approach for ℓ_0-minimization: Application to ARX model segmentation, Automatica 49 (12) (2013) 3646–3653.

[12] O. Burdakov, C. Kanzow, A. Schwartz, On a reformulation of mathematical programs with cardinality constraints, in: D. Gao, N. Ruan, W. Xing (Eds.), Advances in Global Optimization, Vol. 95 of Springer Proceedings in Mathematics & Statistics, Springer International Publishing, 2015, pp. 3–14.

[13] I. Markovsky, How effective is the nuclear norm heuristic in solving data approximation problems?, in: Proc. of the 16th IFAC Symposium on System Identification, Brussels, 2012, pp. 316–321.

[14] M. C. Grant, S. P. Boyd, Graph implementations for nonsmooth convex programs, in: Recent advances in learning and control, Springer, 2008, pp. 95–110.

[15] R. A. Delgado, J. C. Agüero, G. C. Goodwin, A rank-constrained optimization approach: Application to factor analysis, in: 19th IFAC World Congress, Cape Town, South Africa, 2014.

[16] R. P. Aguilera, R. A. Delgado, D. Dolz, J. C. Agüero, Quadratic MPC with ℓ_0-input constraint, in: 19th IFAC World Congress, Cape Town, South Africa, 2014.

[17] R. A. Delgado, J. C. Agüero, G. C. Goodwin, E. M. A. M. Mendes, Application of rank-constrained optimisation to nonlinear system identification, in: 1st IFAC Conference on Modelling, Identification and Control of Nonlinear Systems (MICNON 2015), Saint-Petersburg, Russia, 2015.