

Low-Rank Optimization with Convex Constraints

Christian Grussler, Anders Rantzer, and Pontus Giselsson.

Abstract

The problem of low-rank approximation with convex constraints, which appears in data analysis, system identification, model order reduction, low-order controller design and low-complexity modelling is considered. Given a matrix, the objective is to find a low-rank approximation that meets rank and convex constraints, while minimizing the distance to the matrix in the squared Frobenius norm. In many situations, this non-convex problem is convexified by nuclear norm regularization. However, we will see that the approximations obtained by this method may be far from optimal. In this paper, we propose an alternative convex relaxation that uses the convex envelope of the squared Frobenius norm and the rank constraint. With this approach, easily verifiable conditions are obtained under which the solutions to the convex relaxation and the original non-convex problem coincide. An SDP representation of the convex envelope is derived, which allows us to apply this approach to several known problems. Our example on optimal low-rank Hankel approximation/model reduction illustrates that the proposed convex relaxation performs consistently better than nuclear norm regularization and may outperform balanced truncation.

Index Terms

Low-rank Approximation, Model Reduction, System Identification, k -support norm, Compressed Sensing.

I. INTRODUCTION

Optimization problems with a low-rank (sparsity) constraint have received considerable attention in data driven areas such as image analysis, multivariate linear regression and matrix completion (see, e.g. [1]–[5]), as well as many control subjects such as model order reduction, low order/sparse controller design, low complexity modelling, system identification, etc. [6]–[17]. This is because low-rank approximations allow us to study high dimensional (complex) problems in lower dimensional (simpler) domains. For example, the low-rank approximation of a Hankel operator or matrix requires a smaller number of equations to describe a dynamical system or controller, see, e.g. [6], [8], [18], [19].

For unitarily invariant norms an optimal low-rank approximation can be found by performing a singular value decomposition (SVD). Unfortunately, these approximations usually do not fulfil desired structural constraints such as element-wise nonnegativity, Hankel structure or prescribed entries [2], [3], [12], [20]. Only in a few cases, an explicit solution to the constrained low-rank approximation problem is known [3], [18], [21]. For this reason, other concepts based on convex optimization have been developed [2], [7], [22]–[24]. Many of them rely on nuclear norm regularization, which for particular constraints and assumptions can guarantee a minimum rank solution [2], [25]. As a result, this technique (see [7], [13], [14], [26]) and its extensions (see [16], [17]) has become a standard

tool within control. Nevertheless, it is demonstrated here that nuclear regularization may be far from obtaining the optimal solution to the underlying non-convex problem.

In this work, we study the optimal Frobenius norm low-rank approximation problem with a prescribed target rank and convex constraints (see [Problem 1](#)). We provide an expression for the convex envelope (or equivalently the bi-conjugate) of

$$f(M) = \|N - M\|_F^2 + \chi_{\text{rank}(M) \leq r}(M),$$

where N is a known data matrix and $\chi_{\text{rank}(M) \leq r}(M)$ is the indicator function that allows for matrices of rank at most r . This is used to extend our work in [\[27\]](#) to a more general setting and to provide further analysis.

One formulation of this convex envelope has recently been presented in [\[22\]](#). In this work, we show how the bi-conjugate can be expressed very neatly in terms of the dual norm of the r -norm (the ℓ_2 norm of the r largest singular values). This dual norm is referred to as the r^* norm. A convex relaxation to problems involving f with an additional constraint then naturally arises from the convex envelope of f . We provide guarantees and an example for when a globally optimal solution to our non-convex problem involving f can be found by the proposed convex relaxation. We also show how to construct r^* norms for non-integer valued r . This gives rise to other convex relaxations in which the r can be used as a regularization parameter to trade-off rank and data misfit in the solution.

Further, an SDP-representation of the convex envelope is presented, which allows us to compute solutions to problems with SDP-representable constraints. This is particularly useful if the problem is of medium size (see e.g. [\[8\]](#)), but where it may be tedious to handle a large number of constraints with first order methods [\[28\]](#). Nevertheless, there are several important cases, e.g. Hankel structure, where first order methods can be used to solve problems of large size (see [\[29\]–\[31\]](#) and [\[32\]](#) for available implementations).

The paper is organized as follows. In [Section II](#), we introduce some definitions, recap the unconstrained low-rank approximation problem and define our main problem. Our main approach is derived and discussed in [Section III](#). Extensions of our approach to non-integer valued r are discussed in [Section IV](#) and corresponding SDP-representations are derived in [Section V](#). In [Section VI](#), an application to the open problem of Hankel structure optimal low-rank approximation [\[33\]](#) is presented. These approximations are used to construct reduced order models and to compare their performance with balanced truncation [\[21\]](#). Finally, we draw conclusions and discuss future research in [Section VII](#).

II. BACKGROUND

A. Notations

The following notations for real matrices $X = (x_{ij}) \in \mathbb{R}^{n \times m}$ is used throughout this paper. Without loss of generality, it is assumed that $n \leq m$. Submatrices of X are denoted by

$$X_{(p;q,s;t)} := (x_{ij})_{p \leq i \leq q, s \leq j \leq t} \in \mathbb{R}^{p-q+1 \times s-t+1}.$$

If $X = X^T$ is positive definite (semi-definite) we use the notation $X \succ 0$ ($X \succeq 0$). We also use these notations to describe the relation between two matrices, e.g. $A \succeq B$ means $A - B \succeq 0$.

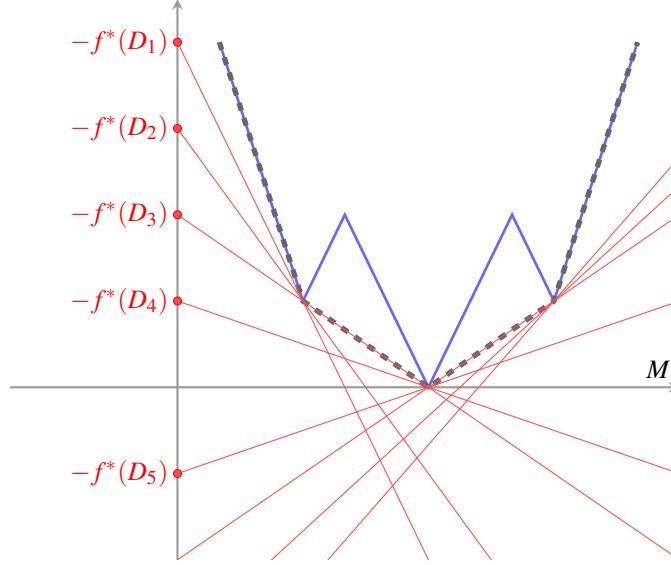


Fig. 1: Schematic plot of — $f(M)$, - - - $f^{**}(M)$ and — tangents through $-f^*(D_i)$.

The non-increasingly ordered singular values of $X \in \mathbb{R}^{n \times m}$ are denoted by $\sigma_1(X) \geq \dots \geq \sigma_n(X)$, counted with multiplicity. The Frobenius inner-product for $X, Y \in \mathbb{R}^{n \times m}$ is defined as

$$\langle X, Y \rangle := \sum_{i=1}^m \sum_{j=1}^n x_{ij} y_{ij} = \text{trace}(X^T Y).$$

Correspondingly, the Frobenius norm is given by

$$\|X\|_F := \sqrt{\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2} = \sqrt{\sum_{i=1}^m \sigma_i^2(X)}.$$

The Frobenius norm is so-called unitarily invariant, i.e. $\|UXV\|_F = \|X\|_F$ for all unitary matrices U and V . The *pseudo-inverse* of X is denoted by X^\dagger (see e.g. [34]).

For a function $f: \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ that is linearly minorized, i.e. there exists $X \in \mathbb{R}^{n \times m}$ with $f(M) \geq \langle M, X \rangle$ for all $M \in \mathbb{R}^{n \times m}$, the *conjugate function* f^* is defined as

$$f^*(D) := \sup_{M \in \mathbb{R}^{n \times m}} [\langle D, M \rangle - f(M)]$$

for all $D \in \mathbb{R}^{n \times m}$. The *bi-conjugate function* of f is given by $f^{**} := (f^*)^*$. It is well-known that f^* and f^{**} are convex (see [35]). Moreover, $f(M) \geq f^{**}(M)$ for all $M \in \mathbb{R}^{n \times m}$. In fact, f^{**} is the largest convex minorizer of f (see [36, Theorem X.1.3.5]), because it is the point-wise supremum of all affine functions majorized by f (see Figure 1).

Finally, if $S \subset \mathbb{R}^{n \times m}$ and $f: \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$, then $\text{argmin}_S f$ denotes the *set of minimizers* of f over S . We write $x^* = \text{argmin}_S f$, if $\text{argmin}_S f = \{x^*\}$ is a singleton. Further, we use $\text{conv}(S)$ to denote the *convex hull* of S .

B. Problem

Let us turn to the underlying problem of this work. We start with the traditional optimal low-rank approximation problem in $\mathbb{R}^{n \times m}$, which is formulated as follows. Given $N \in \mathbb{R}^{n \times m}$ and $r \in \{1, \dots, n\}$, find a solution $M^* \in \mathbb{R}^{n \times m}$ to

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|N - M\|_F^2 \\ & \text{subject to} && \text{rank}(M) \leq r \end{aligned} \tag{1}$$

In case of the Hilbert-Schmidt norm, the natural operator generalization of the Frobenius-norm, this problem has been solved by Schmidt (see [37]). The result is stated next.

Proposition 1. *Let $N \in \mathbb{R}^{n \times m}$ and $r \in \{1, \dots, n\}$. Then,*

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \|N - M\|_F = \|\text{diag}(\sigma_{r+1}(N), \dots, \sigma_n(N))\|_F.$$

All solutions to (1) are given by

$$\text{svd}_r(N) := \left\{ \sum_{i=1}^r \sigma_i(N) u_i v_i^\top : N = \sum_{i=1}^n \sigma_i(N) u_i v_i^\top \text{ is SVD of } N \right\},$$

and each element in $\text{svd}_r(N)$ is referred to as a *standard SVD-approximation* of N . If $\sigma_r(N) = \sigma_{r+1}(N)$, then $\text{svd}_r(N)$ contains infinitely many such solutions, because $\{u_r, u_{r+1}\}$ and $\{v_r, v_{r+1}\}$ are not uniquely determined. Otherwise, $\text{svd}_r(N)$ is a singleton, i.e., if $\sigma_r(N) \neq \sigma_{r+1}(N)$ or $\sigma_r(N) = 0$, and we simply write $\text{svd}_r(N)$ for the unique solution to (1).

This work addresses the following extension of (1).

Problem 1. *Given $N \in \mathbb{R}^{n \times m}$, find $M^* \in \mathbb{R}^{n \times m}$ with $\text{rank}(M^*) \leq r$ such that*

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|N - M\|_F^2 + g(M) \right] = \frac{1}{2} \|N - M^*\|_F^2 + g(M^*),$$

where $g: \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ is a given closed, proper and convex function, i.e., the epi-graph of g is closed, non-empty and convex, respectively.

Compared to (1), **Problem 1** has an additional function g that can be used to add information about the desired solution. Both problems are non-convex due to the rank constraint. Nevertheless, we will see in **Section III** that they can often be solved by convex optimization and semi-definite programming.

In the following, we often use $g(M) \equiv \chi_{\mathcal{C}}(M)$, where

$$\chi_{\mathcal{C}}(M) := \begin{cases} 0, & M \in \mathcal{C} \\ \infty, & M \notin \mathcal{C} \end{cases}$$

is defined to be the indicator function of a (convex) set $\mathcal{C} \subset \mathbb{R}^{n \times m}$. We also use $\chi_{\text{rank}(M) \leq r}$ to denote the indicator function of the set of matrices with at most rank r . In the remainder of this paper, it is assumed that $g + \chi_{\text{rank}(M) \leq r}$ is proper.

C. Nuclear Norm Regularization

One of the most widely used methods to convexify rank constrained problems is to use nuclear norm regularization. It borrows techniques from sparse regularized regression (see [1]), where the ℓ_1 norm is used as a sparsifier.

In our case, rather than having a sparse solution, we are interested in having a small number of non-zero singular values. The nuclear norm imposes an ℓ_1 norm penalty on the singular values. Therefore, for given $N \in \mathbb{R}^{n \times m}$, a matrix version for convexifying [Problem 1](#) reads

$$\min_{M \in \mathbb{R}^{n \times m}} \frac{1}{2} \|N - M\|_F^2 + \mu \|M\|_{1*} + g(M), \quad (2)$$

where $g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ is a closed and proper convex function. The simplicity of this convexification, as well as the results in [2], [7], stimulated a large growth in the application of this method. However, it is often challenging to choose μ a priori in order to obtain a solution of specific rank. Commonly one assumes that the rank as a function of μ looks like a staircase, i.e., a large/small μ decreases/increases the rank.

In general, this heuristic does not return an optimal solution to [Problem 1](#). In particular, in the case $g = 0$, one usually cannot choose μ such that the SVD-approximation is obtained. Finally, there is no certificate for checking whether a solution is a minimizer of [Problem 1](#).

III. THE r^* APPROACH

In the following, we consider the problem of finding solutions to [Problem 1](#). Our approach is based on convex relaxations of [Problem 1](#) by means of what we call the r^* norms. These norms are defined in the following lemma.

Lemma 1. *Let $M \in \mathbb{R}^{n \times m}$, and $r \in \{1, \dots, n\}$. Then,*

$$\|M\|_r := \sqrt{\sum_{i=1}^r \sigma_i^2(M)} = \sup_{\substack{\|X\|_F=1 \\ \text{rank}(X) \leq r}} \langle M, X \rangle \quad (3)$$

is a unitarily invariant norm with dual norm

$$\|M\|_{r^*} := \max_{\|X\|_r \leq 1} \langle M, X \rangle = \max_{\sum_{i=1}^r s_i^2 \leq 1} \left[\sum_{i=1}^r \sigma_i(M) s_i + s_r \sum_{i=r+1}^n \sigma_i(M) \right].$$

Moreover,

$$\|M\|_1 \leq \dots \leq \|M\|_n = \|M\|_F = \|M\|_{n^*} \leq \dots \leq \|M\|_{1^*}, \quad (4)$$

$$\text{rank}(M) \leq r \text{ if and only if } \|M\|_r = \|M\|_F = \|M\|_{r^*}. \quad (5)$$

A proof to this lemma is provided in [Appendix B](#). Notice that $\|M\|_1 = \sigma_1(M)$ is equal to the spectral norm and its dual norm $\|M\|_{1^*} = \sum_{i=1}^n \sigma_i(M)$ is equal to the nuclear (trace norm). These norms can be formulated using convex linear matrix inequalities (see [2], [7]). In [Section III](#) it is shown that the same holds true for $\|\cdot\|_r^2$ and $\|\cdot\|_{r^*}^2$.

Next we show that the r^* norm can be used to construct the largest convex minorizer (convex envelope) of

$$f(M) := \frac{1}{2} \|N - M\|_F^2 + \chi_{\text{rank}(M) \leq r}(M).$$

Theorem 1. Let $N \in \mathbb{R}^{n \times m}$, and $r \in \{1, \dots, n\}$. Then the conjugate and bi-conjugate functions of

$$f(M) := \frac{1}{2} \|N - M\|_F^2 + \chi_{\text{rank}(M) \leq r}(M)$$

are given by

$$f^*(D) = \frac{1}{2} \|N + D\|_r^2 - \frac{1}{2} \|N\|_F^2, \quad (6)$$

$$f^{**}(M) = \frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 \quad (7)$$

for all $D, M \in \mathbb{R}^{n \times m}$.

A proof to [Theorem 1](#) can be found in [Appendix C](#). Note that by Fenchel duality (see [[38](#), Section 31]) the following Lemma holds.

Lemma 2. Let $f, g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ be such that g is proper, closed and convex. Then,

$$\inf_{M \in \mathbb{R}^{n \times m}} [f(M) + g(M)] \geq - \inf_{D \in \mathbb{R}^{n \times m}} [f^*(D) + g^*(-D)] \quad (8)$$

$$= \inf_{M \in \mathbb{R}^{n \times m}} [f^{**}(M) + g(M)]. \quad (9)$$

If M^* is a solution to [\(9\)](#) such that $f(M^*) = f^{**}(M^*)$, then M^* is also a solution to the left-hand side of [\(8\)](#).

Therefore, we can construct the dual and bi-dual problems to [Problem 1](#) as

$$- \min_{D \in \mathbb{R}^{n \times m}} \left[g^*(-D) + \frac{1}{2} \|N + D\|_r^2 - \frac{1}{2} \|N\|_F^2 \right], \quad (A)$$

$$\min_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 + g(M) \right], \quad (B)$$

which are accompanied by the next central result.

Proposition 2. Let $N \in \mathbb{R}^{n \times m}$ and $g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ be a closed proper convex function. Then for all $r \in \{1, \dots, n\}$

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|N - M\|_F^2 + g(M) \right] \geq - \min_{D \in \mathbb{R}^{n \times m}} \left[g^*(-D) + \frac{1}{2} \|N + D\|_r^2 - \frac{1}{2} \|N\|_F^2 \right] \quad (C)$$

$$= \min_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 + g(M) \right].$$

Assume that [\(B\)](#) has a minimizer M^* with $\text{rank}(M^*) \leq r$. Then,

$$\underset{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}}{\text{argmin}} \left[\frac{1}{2} \|N - M\|_F^2 + g(M) \right] \subset \underset{M \in \mathbb{R}^{n \times m}}{\text{argmin}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 + g(M) \right].$$

Thus obtaining a rank- r solution to the convex relaxation problem [\(B\)](#) implies solving the original non-convex problem. This is why we suggest to use [\(B\)](#) instead of the the nuclear norm heuristic (see [\(2\)](#) in [Section II-C](#)) as convex relaxation to [Problem 1](#). Nevertheless, in general there may be a duality-gap for some choices of g (see [Section V](#)). This is reflected by the inequality in [\(C\)](#). Fortunately, there are many situations with no duality-gap. Next, an important case is discussed to provide additional insights.

Proposition 3. Assume that D^* is a solution to (A) and $\sigma_r(N+D^*) \neq \sigma_{r+1}(N+D^*)$ or $\sigma_r(N+D^*) = 0$. Then there is no duality gap in (C) and $\text{svd}_r(N+D^*)$ is the unique minimizing argument of Problem 1, i.e.

$$\text{svd}_r(N+D^*) = \underset{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}}{\text{argmin}} \left[\frac{1}{2} \|N - M\|_F^2 + g(M) \right].$$

Proposition 3 provides a simple sufficient condition for the uniqueness of a solution to Problem 1, which in many applications is fulfilled (see Section VI). However, this is not a necessary condition. A proof of Proposition 3 is given in a more general setting in Theorem 2, which also allows us to say something about the rank of the solution to the convex relaxation if there is a duality-gap.

Theorem 2. Let D^* and M^* be solutions to (A) and (B), respectively. Further, suppose that an SVD of $N+D^*$ is given by $N+D^* = \sum_{i=1}^n \sigma_i u_i v_i^\top$ with $\sigma_r = \dots = \sigma_{r+s} \neq \sigma_{r+s+1}$, where $s = n - r$ if $\sigma_n = \sigma_r$. Then,

$$M^* \in \text{conv}(\text{svd}_r(N+D^*)).$$

In particular, $\text{rank}(M^*) \leq r + s$. Moreover, if $\sigma_r \neq \sigma_{r+1}$ or $\sigma_r = 0$, then $M^* = \text{svd}_r(N+D^*)$.

A proof to this theorem is given in Appendix D. Observe that whenever (B) does not have a unique solution, it follows by Theorem 2 that

$$\sigma_r(N+D^*) = \sigma_{r+1}(N+D^*)$$

for all solutions D^* to (A). Furthermore, Theorem 2 shows that $\text{svd}_r(N)$ with $\sigma_r(N) \neq \sigma_{r+1}(N)$ can be determined by solving a convex problem.

Corollary 1. Let $N \in \mathbb{R}^{n \times m}$, and $r \in \{1, \dots, n\}$. Then,

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \frac{1}{2} \|N - M\|_F^2 = \frac{1}{2} \|N\|_F^2 - \frac{1}{2} \|N\|_r^2 = \min_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 \right]$$

and

$$\text{svd}_r(N) \subset \underset{M \in \mathbb{R}^{n \times m}}{\text{argmin}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle \right].$$

If $\sigma_r(N) \neq \sigma_{r+1}(N)$ or $\sigma_r = 0$ then

$$\text{svd}_r(N) = \underset{M \in \mathbb{R}^{n \times m}}{\text{argmin}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle \right].$$

Proof. Since $g = 0$, $g^*(D)$ is finite if and only if $D = 0$. Thus the result follows by Theorem 2. \square

The low-rank inducing property of the r^* norm can also be seen by characterizing the extreme points of its unit ball.

Lemma 3. The set of the extreme points of the unit-ball $B_1 := \{X : \|X\|_{r^*} \leq 1\}$ is

$$E := \{X \in \mathbb{R}^{n \times m} : \|X\|_F = 1, \text{rank}(X) \leq r\}.$$

Hence, $B_1 = \text{conv}(E)$.

Proof. By (3) in Lemma 1, it holds that for all $N \in \mathbb{R}^{n \times m}$

$$\sup_{M \in \text{conv}(E)} \langle N, M \rangle = \|N\|_r = \sup_{M \in B_1} \langle N, M \rangle. \quad (10)$$

Since $\text{conv}(E)$ and B_1 are closed convex sets, [38, Corollary 13.1.1.] implies that $B_1 = \text{conv}(E)$. If a point $\bar{M} \in E$ is not an extreme point of E , then

$$\bar{M} = \sum_i \alpha_i M_i, \quad \text{with} \quad \sum_i \alpha_i = 1,$$

such that

$$M_i \in K \setminus \{\bar{M}\} \quad \text{and} \quad \alpha_i > 0 \quad \text{for all } i.$$

Hence, by the Cauchy-Schwarz inequality we conclude that

$$1 = \langle \bar{M}, \bar{M} \rangle = \sum_i \alpha_i \langle \bar{M}, M_i \rangle \leq \sum_i \alpha_i = 1.$$

However, this can only be true if $\langle \bar{M}, M_i \rangle = 1$ for all i . Equivalently, $\bar{M} = M_i$ and that is a contradiction. \square

Finally, the preceding results cover several extensions of Problem 1. By letting $N = \text{diag}(v)$ and $M = \text{diag}(w)$ for $v, w \in \mathbb{R}^n$, there are analogous norms for vector-valued problems, (see e.g. [39], [40]) where our analysis carries over. Further, it is possible to consider the weighted case

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|W(N - M)\|_F^2 + g(M) \right], \quad (11)$$

where $W \in \mathbb{R}^{l \times n}$ and $\text{rank}(W) = n$. Since $\text{rank}(\tilde{M}) = \text{rank}(W^\dagger \tilde{M}) = \text{rank}(M)$, (11) can be reformulated such that it fits Problem 1 by letting $\tilde{g}(\tilde{M}) := g(W^\dagger \tilde{M})$:

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|W(N - M)\|_F^2 + g(M) \right] = \min_{\substack{\tilde{M} \in \mathbb{R}^{l \times m} \\ \text{rank}(\tilde{M}) \leq r}} \left[\frac{1}{2} \|WN - \tilde{M}\|_F^2 + \tilde{g}(\tilde{M}) \right].$$

Since another inner product and norm is defined by W as

$$\begin{aligned} \|W(N - M)\|_F^2 &= \text{trace}((N - M)^\top W^\top W(N - M)) \\ &=: \langle N - M, N - M \rangle_{W^\top W}, \end{aligned}$$

a suitable W may enable us to satisfy the requirements of Proposition 3 in situations where the Frobenius norm fails. In particular, W may be used for iterative re-weighting. For vector-valued problems, this generalizes the idea of ℓ_1 norm re-weighting (see [41]) to r^* norms.

IV. REAL-VALUED EXTENSION

In the following, it is shown that allowing r to be real-valued can be considered as a regularization parameter. Unlike typical regularization methods (see [7], [22]), this parameter has a close relationship to the rank of the corresponding solutions.

It suffices to discuss the case where Proposition 3 does not apply. Therefore, let

$$D_t^* := \underset{D \in \mathbb{R}^{n \times m}}{\text{argmin}} \left[g^*(-D) + \frac{1}{2} \|N + D\|_t^2 \right],$$

and

$$M_t^* := \operatorname{argmin}_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{t^*}^2 - \langle N, M \rangle + g(M) \right].$$

be defined for all $t \in \{1, \dots, n\}$, and assume that there exists $r \in \mathbb{N}$ with

$$\sigma_r(N + D_r^*) = \sigma_{r+1}(N + D_r^*) \quad \text{and} \quad \operatorname{rank}(M_r^*) > r.$$

Furthermore, let

$$\frac{1}{2} \|N - M_r^*\|_F^2 + g(M_r^*) > \frac{1}{2} \|N - M_{r+1}^*\|_F^2 + g(M_{r+1}^*)$$

with

$$\operatorname{rank}(M_{r+1}^*) > \operatorname{rank}(M_r^*).$$

In such a scenario, one often faces the situation that $\operatorname{rank}(M_r^*)$ is small, but the cost $\frac{1}{2} \|N - M_r^*\|_F^2 + g(M_r^*)$ is poor, whereas $\frac{1}{2} \|N - M_{r+1}^*\|_F^2 + g(M_{r+1}^*)$ may be acceptable, but $\operatorname{rank}(M_{r+1}^*)$ is too large. Then a trade-off between M_r^* and M_{r+1}^* is desired. Such a trade-off can be achieved by letting r become non-integer valued in the r norm. The r norm extends to

$$\|\cdot\|_r := \sqrt{\sum_{i=1}^{\lfloor r \rfloor} \sigma_i^2(\cdot) + (r - \lfloor r \rfloor) \sigma_{\lfloor r \rfloor}^2(\cdot)}, \quad (12)$$

where $\lfloor r \rfloor := \max\{z \in \mathbb{Z} : z \leq r\}$ and $\lceil r \rceil := \min\{z \in \mathbb{Z} : z \geq r\}$. For $r \in \mathbb{N}$ and $\alpha \in [0, 1]$ we have

$$\|\cdot\|_{r+\alpha}^2 = (1 - \alpha) \|\cdot\|_r^2 + \alpha \|\cdot\|_{r+1}^2, \quad (13)$$

which means that $\|\cdot\|_{r+\alpha}^2$ is a convex combination of $\|\cdot\|_r^2$ and $\|\cdot\|_{r+1}^2$, and thus indicates its usefulness in supplying the desired trade-off solution. Similar to [Theorem 2](#), it remains true by [Proposition 4](#) that $\operatorname{rank}(M_r^*) \leq \lceil r \rceil + s$ if $r \in \mathbb{R}_{\geq 1}$ and

$$\sigma_{\lceil r \rceil}(N + D_r^*) = \dots = \sigma_{\lceil r \rceil + s}(N + D_r^*) > \sigma_{\lceil r \rceil + s + 1}(N + D_r^*). \quad (14)$$

Hence, allowing r to assume real values may allow us to find solutions of both lower rank and lower cost. Next we look at the dependency of s on r in [\(14\)](#). We define

$$F(D, r) := g^*(-D) + \frac{1}{2} \|N + D\|_r^2 + \frac{1}{2} \|N\|_F^2.$$

Using the piecewise linearity in [\(13\)](#), it can be shown that F is (jointly) continuous on the relative interior of its domain. Therefore, Berge's Maximum Theorem (see [\[42, p. 116\]](#)) implies that the parameter depending set

$$\mathcal{C}^*(r) := \operatorname{argmin}_{D \in \mathbb{R}^{n \times m}} \left[g^*(-D) + \frac{1}{2} \|N + D\|_r^2 + \frac{1}{2} \|N\|_F^2 \right]$$

is upper hemicontinuous in r . This means that for all $r \in [1, \min\{m, n\}]$ and all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $t \geq 1$

$$|t - r| < \delta \Rightarrow \mathcal{C}^*(t) \subset \mathcal{B}_\varepsilon(\mathcal{C}^*(r)), \quad (15)$$

where

$$\mathcal{B}_\varepsilon(\mathcal{C}^*(r)) := \{X \in \mathbb{R}^{n \times m} : \exists D \in \mathcal{C}^*(r) \text{ with } \|X - D\|_F < \varepsilon\}.$$

For simplicity assume that D_r^* is unique. By (15) and the continuity of the singular values (see [37, Corollary 4.9]), it follows that a sufficiently small increase of r does not increase s in (14). Hence, just as for nuclear norm regularization, $\text{rank}(M_t^*)$ often looks like a staircase as t varies over $[r, r+1]$ (see Figure 2b in Section VI). In summary, real-valued r can be considered as a regularization parameter, similar to other regularization methods such as in [7], [22].

V. SDP-REPRESENTATIONS

Next we develop SDP-representations of the problems (A) and (B) under the assumption that g is SDP-representable. We start with an SDP-representation of the optimization problem

$$\min_{D \in \mathbb{R}^{n \times m}} \|N+D\|_r^2, \quad (16)$$

where $\|\cdot\|_r$ is defined as in (12) and $r \in [1, n]$. Let $T \in \mathbb{R}^{n \times n}$ be such that

$$T \succeq (N+D)(N+D)^\top.$$

Then $\sigma_i(T) \geq \sigma_i^2(N+D)$ for all i such that $1 \leq i \leq n$ (see [34, Corollary 7.7.4]) and $\text{trace}(T) = \sum_{i=1}^n \sigma_i(T)$. Hence,

$$\|N+D\|_r^2 \leq \text{trace}(T) - (n-r)\sigma_n(T),$$

which implies that

$$\|N+D\|_r^2 \leq \min_{T \succeq (N+D)(N+D)^\top} \text{trace}(T) - (n-r)\sigma_n(T). \quad (17)$$

In particular, equality in (17) can be achieved with

$$T^* := \sum_{i=1}^{\lceil r \rceil} \sigma_i^2(N+D) u_i u_i^\top + \sigma_{\lceil r \rceil}^2(N+D) \sum_{i=\lceil r \rceil+1}^n u_i u_i^\top,$$

where $N+D = \sum_{i=1}^n \sigma_i(N+D) u_i v_i^\top$ is an SVD of $N+D$. Using the Schur-complement condition for $T - (N+D)(N+D)^\top \succeq 0$ (see [34, Theorem 7.7.7]) yields that

$$\begin{aligned} & \underset{D, T, \gamma}{\text{minimize}} && \text{trace}(T) - \gamma(n-r) \\ & \text{subject to} && \begin{pmatrix} T & N+D \\ (N+D)^\top & I \end{pmatrix} \succeq 0, \quad T \succeq \gamma I, \quad D \in \mathbb{R}^{n \times m}. \end{aligned}$$

is an SDP-representation for (16). Then, an SDP-formulation of (B) can be obtained by deriving the dual of this optimization problem as

$$\begin{aligned} & \underset{M, P, W}{\text{minimize}} && \frac{1}{2} \text{trace}(W) - \text{trace}(N^\top M) + g(M) \\ & \text{subject to} && \begin{pmatrix} I-P & M \\ M^\top & W \end{pmatrix} \succeq 0, \quad P \succeq 0, \quad \text{trace}(P) = n-r. \end{aligned} \quad (18)$$

VI. MODEL ORDER REDUCTION

In system and control, the rank of a Hankel matrix/operator is important, because it determines the order, e.g. of a linear time invariant discrete-time system

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k, \\ y_k &= Cx_k + Du_k, \end{aligned} \tag{19}$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n_u}$, $C \in \mathbb{R}^{n_y \times n}$ and $D \in \mathbb{R}^{n_y \times n_u}$. Note that if (A, B, C, D) is a minimal realization, then n is the order of the system and thus decides how costly it is to simulate or control the system (see, e.g. [8], [19], [21]). As a result, the field of model order reduction has emerged [21]. Whereas the Adamyan-Arov-Krein theorem [21] answers the question of optimal low-rank approximation of infinite dimensional Hankel operators, the finite dimensional case

$$\begin{aligned} &\underset{M}{\text{minimize}} \quad \|N - M\|_F^2 \\ &\text{subject to} \quad M \in \mathcal{H}, \text{rank}(M) \leq r, \end{aligned} \tag{20}$$

where $N \in \mathcal{H} := \{H : H \text{ is Hankel}\}$, is still an open problem [33]. The finite dimensional case (20) is important, e.g. for model approximation or system identification (see [12], [19]), where N is formed through the known or measured impulse response, $h_0 = D$, $h_t = CA^{t-1}B$, $t \geq 1$, of a stable linear system (19):

$$N = H_{k,l+1} := \begin{pmatrix} h_1 & h_2 & \cdots & h_{l+1} \\ h_2 & h_3 & \cdots & h_{l+2} \\ \vdots & \vdots & & \vdots \\ h_k & h_{k+1} & \cdots & h_{k+l} \end{pmatrix}.$$

Assuming that $k, l \geq n$, it holds that $\text{rank}(N) \leq n$ and N can be mapped onto a minimal realization of (19) through Kung's (or Ho-Kalman-Kung) algorithm [12], [19], [21]. Moreover, also a Hankel structured rank- r approximation M^* of N can be mapped by Kung's algorithm onto a linear system $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ of order r if $\text{rank}(M_{(1:n_y, 1:n_u)}^*) = r$. The system matrices are derived as

$$\begin{aligned} \hat{A} &= O^\dagger M_{(1:k, 2:l+1)}^* R^\dagger, \quad \hat{B} = R_{(1:r, 1:n_u)}, \\ \hat{C} &= O_{(1:n_y, 1:r)}, \quad \hat{D} = h_0, \end{aligned} \tag{21}$$

where an SVD of $M_{(1:k, 1:l)}^* = \sum_{i=1}^r \sigma_i u_i v_i^\top$ determines

$$O := \begin{pmatrix} \sigma_1 u_1 & \cdots & \sigma_r u_r \end{pmatrix} \quad \text{and} \quad R := \begin{pmatrix} \sigma_1 v_1 & \cdots & \sigma_r v_r \end{pmatrix}^\top.$$

Consequently, the impulse response matches M^* , i.e. it fulfils

$$\hat{C} \hat{A}^{t-1} \hat{B} = \begin{cases} M_{((t-1)n_y+1:n_y, 1:n_u)}^*, & 1 \leq t \leq k, \\ M_{(1:n_y, (t-1)n_u+1:n_u)}^*, & k+1 \leq t \leq k+l. \end{cases}$$

In the following, we compare the performance of the r^* approach (B) and nuclear norm regularization (2) to balanced truncation for the minimal system (19) of order 10 with

$$A = \text{diag}(0, 0.1, \dots, 0.9), \quad C = B^\top = (1, \dots, 1), \quad D = 0 \tag{22}$$

and $N = H_{71,71}$ being the intrinsic Hankel matrix. We use Kung’s algorithm to map the Hankel matrix approximations of the convex methods onto systems and compare their H_∞ norm errors (see Figure 2a) with balanced truncation. Further, we construct Hankel matrix approximations of N from the balanced truncated models and compare their Frobenius norm errors with those of the convex methods (see Figure 2a).

By the rank evolution in Figure 2b, we can see that the r^* approach exhibits the expected staircase behaviour as discussed in Section IV. Further, it can be observed that there is a zero duality gap for all $r \in \{1, \dots, 9\}$. Thus by Proposition 2, the Frobenius norm error in Figure 2a is the lowest for the r^* approach. In particular, nuclear norm regularization performs 2 – 34 times worse than BT, whereas the r^* approach has about 9 – 16 % smaller error than BT. An even stronger error difference reveals for the corresponding system errors, where our method performs 17 – 39 % better than BT and the nuclear norm 3 – 200 times worse. Finally, the complete evolution of the normalized errors for the r^* approach in Figure 2b shows that good approximations can be achieved for both the system as well as N . The small gap between the two errors, which increases with r , is due to the large sampling horizon in N . The horizon of 141 samples insures that the first 10 singular values in N are close to the Hankel singular values of the system.

Note that a larger horizon would improve the approximation quality even further. However, our chosen horizon seems to give a good trade-off between error performance and computational cost when solving (18) through conventional SDP solvers (see e.g. [43]). Moreover, a smaller horizon seems to mainly affect the quality of higher order approximations. Finally, note that our method does not necessarily need to reduce the original system. It could also be used after an initial reducing step through other methods [21].

An implementation of our example can be found in [32].

VII. CONCLUSION

In this work, a method for determining Frobenius norm optimal low-rank approximations with convex constraints has been studied. The main benefits of our approach are that it is essentially regularization parameter free and may give a certificate of optimality. Moreover, we have seen that our approach can be turned into a regularization dependent method, where, unlike other approaches, the parameter has a direct relationship to the desired rank (see Section IV). The model reduction example shows the superiority of our approach over the nuclear-norm heuristic as well as balanced truncation. In addition, our approach allows us to impose further convex constraints onto the impulse response. In the future, we would like to investigate the distinct properties of such approximations, e.g. error bounds, as well as their effectiveness in system identification problems. Furthermore, it would be interesting to see how system characteristics effect a possible duality gap. Most of our results can be extended to Hilbert-Schmidt operators. In case of Hankel operators, the singular values translate to Hankel singular values. Nevertheless, our results cannot be easily extended to other unitarily invariant norms, e.g. the spectral norm. This is because other norms often lack the following properties: (1) The norm of a difference of two matrices is not decomposable such that the convex envelope can be easily derived (see [30] for more details), (2) If the norm does not depend on all singular values, then even for $g = 0$ there are (infinitely) many solutions and thus minimizing its convex envelope

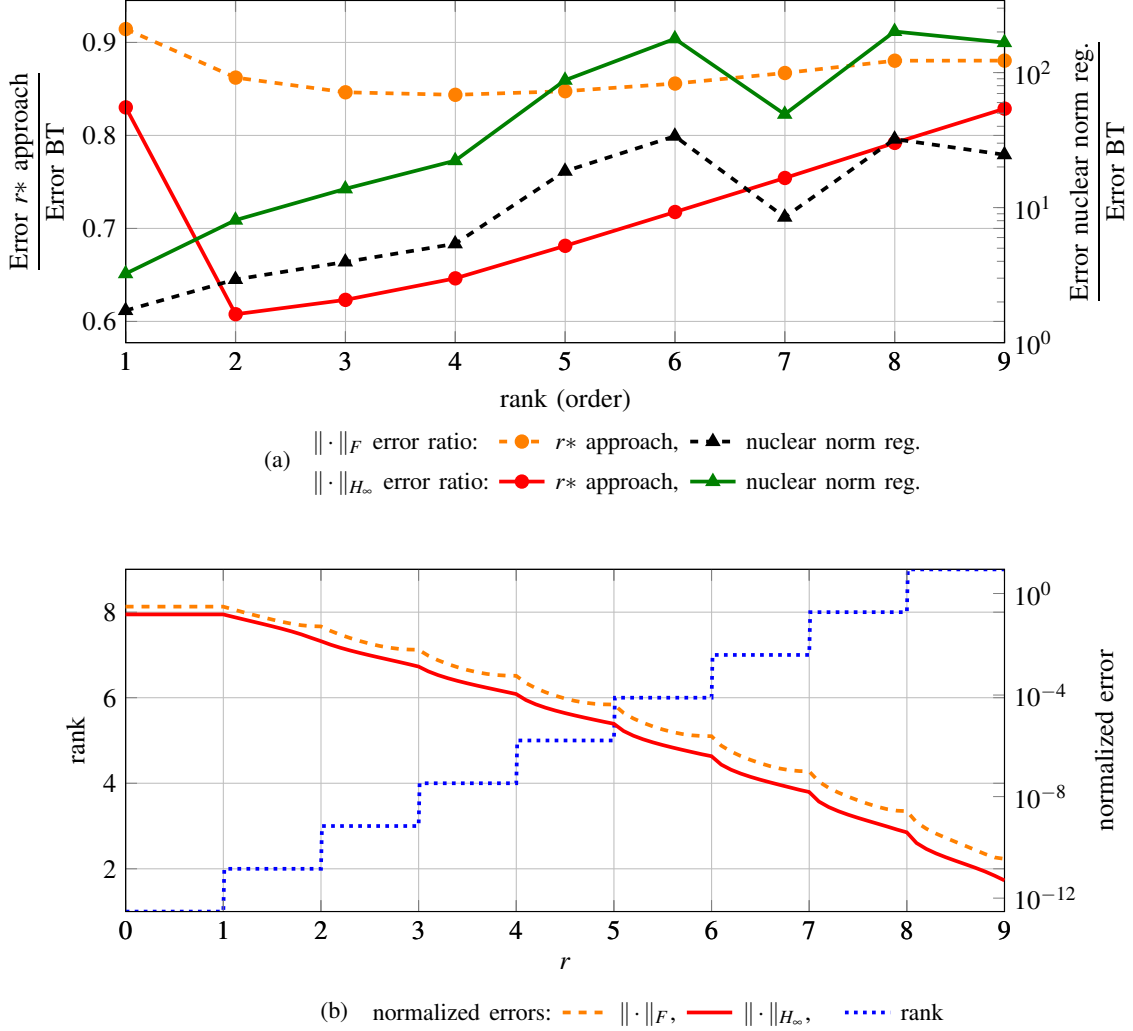


Fig. 2: Model Order Reduction for (22) – (a) Frobenius norm error of the best achievable approximations to $N = H_{71,71}$ in (20) through the r^* approach (B) (left y-axis in linear scale) as well as the nuclear norm regularization (2) (right y-axis in log scale); H_∞ norm error of the corresponding reduced order systems (21) resulting from Kung’s algorithm; both errors are divided by the analogous errors of the balanced truncated (BT) models. (b) Evolution of the rank (left y-axis) as well as the normalized Frobenius and H_∞ norm errors (right y-axis in log scale) of the r^* approach with real-valued $r \in (0, 9]$.

would almost certainly result in a high rank convex combination of these solutions. Similar effects can be expected for $g \neq 0$.

Finally note that our approach can also be used to numerically evaluate the performance of heuristics where no relationships to the optimal solutions are known.

APPENDIX

A. Subdifferentials

Let $f : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ be a convex function, then the subdifferential of f in $X \in \mathbb{R}^{n \times m}$ is defined as

$$\partial f(X) := \{Z : f(Y) \geq f(X) - \langle Y - X, Z \rangle \text{ for all } Y\}.$$

The following proposition on the subgradienital of $\|\cdot\|_r$ has been shown in [44] for $r \in \mathbb{N}$. It is straightforward to extend it to the real-valued case.

Proposition 4. Let $A \in \mathbb{R}^{n \times m} \setminus \{0\}$, $r \in [1, n]$ and $\bar{r} := \lceil r \rceil$. Further, let an SVD of A be given by $A = \sum_{i=1}^n \sigma_i u_i v_i^\top$ with

$$\sigma_{\bar{r}-t} \neq \sigma_{\bar{r}-t+1} = \dots = \sigma_{\bar{r}} = \dots = \sigma_{\bar{r}+s} \neq \sigma_{\bar{r}+s+1},$$

where $t = \bar{r}$ and $s = n - \bar{r}$ if $\sigma_1 = \sigma_{\bar{r}}$ and $\sigma_n = \sigma_{\bar{r}}$, respectively. Then $M \in \partial \|A\|_r$ if and only if

$$M = \frac{1}{\|A\|_r} \left(\sum_{i=1}^{\bar{r}-t} \sigma_i u_i v_i^\top + \sigma_{\bar{r}} R \right),$$

$$R = \begin{pmatrix} u_{\bar{r}-t+1} & \dots & u_{\bar{r}+s} \end{pmatrix} T \begin{pmatrix} v_{\bar{r}-t+1} & \dots & v_{\bar{r}+s} \end{pmatrix}^\top,$$

where $T \succeq 0$, $\|T\|_{1*} = t - \bar{r} + r$, and $\|T\|_1 \leq 1$. Moreover,

$$\partial \|0\|_r = \{M \in \mathbb{R}^{n \times m} : \|M\|_{r*} \leq 1\}.$$

It is readily seen that Proposition 4 is equivalent to

$$\partial \|A\|_{r*} = \frac{1}{\|A\|_r} \text{conv}(\text{svd}_r(A)).$$

B. Proof of Lemma 1

Proof. Let $1 \leq r \leq n$, $M \in \mathbb{R}^{n \times m}$ and the function $g : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ be defined by

$$g(x_1, \dots, x_n) := \|\text{diag}(x_1, \dots, x_n)\|_r.$$

The unitary invariance of $\|\cdot\|_r$ follows by [34, Theorem 7.4.7.2.], because g is a symmetric gauge function. By [34, Corollary 7.4.1.3.] it holds that

$$\sup_{\substack{\|X\|_F=1 \\ \text{rank}(X) \leq r}} \langle X, M \rangle = \sup_{\sum_{i=1}^r \sigma_i^2(X)=1} \sum_{i=1}^r \sigma_i^2(X) \sigma_i(M) = \|M\|_r.$$

Then the r^* -norm inherits the unitary invariance of the r -norm and with $\Sigma := \text{diag}(\sigma_1(M), \dots, \sigma_n(M))$ it follows that

$$\|M\|_{r*} = \|\Sigma\|_{r*} = \max_{\|X\|_r \leq 1} \langle \Sigma, X \rangle = \max_{\sum_{i=1}^r \sigma_i^2(X)=1} \sum_{i=1}^n \sigma_i(M) \sigma_i(X) = \max_{\sum_{i=1}^r \sigma_i^2(X) \leq 1} \left[\sum_{i=1}^r \sigma_i(M) \sigma_i(X) + \sigma_r(X) \sum_{i=r+1}^n \sigma_i(M) \right].$$

The third equality follows by [34, Corollary 7.4.1.3.]. Hence,

$$\|M\|_{1*} = \max_{\sum_{i=1}^n s_i^2=1} \sum_{i=1}^n \sigma_i(M) s_i \geq \dots \geq \max_{\sum_{i=1}^n s_i^2=1} \sum_{i=1}^n \sigma_i(M) s_i = \|M\|_{n*} = \|M\|_F.$$

Moreover, by the definition of the r -norm

$$\|M\|_F = \|M\|_n \geq \cdots \geq \|M\|_1$$

and therefore (4) is shown. In particular,

$$\|M\|_{r^*} = \max_{\sum_{i=1}^r s_i^2 = 1} \sum_{i=1}^n \sigma_i(M) s_i \geq \|M\|_F \geq \max_{\sum_{i=1}^r s_i^2 = 1} \sum_{i=1}^r \sigma_i(M) s_i = \|M\|_r.$$

Obviously, $\|M\|_F = \|M\|_r$ if and only if $\text{rank}(M) \leq r$, and thus $\|M\|_{r^*} = \|M\|_r$ if and only if $\text{rank}(M) \leq r$. \square

C. Proof of Theorem 1

Proof. The conjugate function satisfies

$$\begin{aligned} f^*(D) &= \sup_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\langle D, M \rangle - \frac{1}{2} \|N - M\|_F^2 \right] \\ &= \sup_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} -\frac{1}{2} \|N - M + D\|_F^2 + \langle D, N \rangle + \frac{1}{2} \|D\|_F^2 \\ &= -\frac{1}{2} \|N + D\|_F^2 + \frac{1}{2} \|N + D\|_r^2 + \langle D, N \rangle + \frac{1}{2} \|D\|_F^2 \\ &= -\frac{1}{2} \|N\|_F^2 + \frac{1}{2} \|N + D\|_r^2, \end{aligned}$$

where the third equality follows by Proposition 1. Hence,

$$\begin{aligned} f^{**}(M) &= \sup_{D \in \mathbb{R}^{n \times m}} \left[\langle D, M \rangle + \frac{1}{2} \|N\|_F^2 - \frac{1}{2} \|N + D\|_r^2 \right] \\ &= \sup_{D \in \mathbb{R}^{n \times m}} \left[\langle D - N, M \rangle + \frac{1}{2} \|N\|_F^2 - \frac{1}{2} \|D\|_r^2 \right] \\ &= \frac{1}{2} \|N\|_F^2 - \langle N, M \rangle + \sup_{D \in \mathbb{R}^{n \times m}} \left[\langle D, M \rangle - \frac{1}{2} \|D\|_r^2 \right] \\ &= \frac{1}{2} \|N\|_F^2 - \langle N, M \rangle + \frac{1}{2} \|M\|_{r^*}^2, \end{aligned}$$

where the last equality follows by [38, Corollary 15.3.1] with

$$\frac{1}{2} \|\cdot\|_{r^*}^2 = \left(\frac{1}{2} \|\cdot\|_r \right)^*.$$

\square

D. Proof of Theorem 2

Proof. If D^* and M^* are solutions to (A) and (B), respectively, then by [38, Theorem 31.1] it holds that

$$f^{**}(M^*) = \langle D^*, M^* \rangle - f^*(D^*),$$

where f^* and f^{**} are given by (6) and (7). Hence, by [38, Theorem 23.5.] it follows that

$$M^* \in \partial_D \frac{1}{2} \|N + D\|_r^2 \Big|_{D=D^*} = \|N + D^*\|_r \partial_D \|N + D\|_r \Big|_{D=D^*}$$

and invoking Proposition 4 proves the result. \square

ACKNOWLEDGMENT

The authors would like to thank Andrey Ghulchak for his useful comments and numerous counter-examples. All authors are members of the LCCC Linnaeus Center and the eLLIIT Excellence Center at Lund University. The first author is financially supported by the Swedish Research Council through the project 621-2012-5357. The first and third authors are financially supported by the Swedish Foundation for Strategic Research.

REFERENCES

- [1] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, 2015.
- [2] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [3] G. C. Reinsel and R. Velu, *Multivariate Reduced-Rank Regression: Theory and Applications*, ser. Lecture Notes in Statistics. Springer New York, 1998, vol. 136.
- [4] R. Vidal, Y. Ma, and S. S. Sastry, *Generalized Principal Component Analysis*, ser. Interdisciplinary Applied Mathematics. Springer-Verlag New York, 2016, vol. 40.
- [5] L. Eldén, *Matrix methods in data mining and pattern recognition*. SIAM, 2007.
- [6] A. C. Antoulas, “On the approximation of hankel matrices,” in *Operators, Systems and Linear Algebra: Three Decades of Algebraic Systems Theory*, U. Helmke, Ed. Vieweg+Teubner Verlag, 2013, pp. 17–22.
- [7] M. Fazel, H. Hindi, and S. P. Boyd, “A rank minimization heuristic with application to minimum order system approximation,” in *Proceedings of the 2001 American Control Conference.*, vol. 6, 2001, pp. 4734–4739.
- [8] D. Ankelhed, “On design of low order H-infinity controllers,” Ph.D. dissertation, Linköping University, 2011.
- [9] D. M. Zoltowski, N. Dhingra, F. Lin, and M. R. Jovanović, “Sparsity-promoting optimal control of spatially-invariant systems,” in *2014 American Control Conference*, June 2014, pp. 1255–1260.
- [10] A. Zare, Y. Chen, M. R. Jovanović, and T. T. Georgiou, “Low-complexity modeling of partially available second-order statistics: Theory and an efficient matrix completion algorithm,” *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1368–1383, March 2017.
- [11] C. Grussler, A. Zare, M. R. Jovanovic, and A. Rantzer, “The use of the r^* heuristic in covariance completion problems,” in *55th IEEE Conference on Decision and Control (CDC)*, 2016.
- [12] D. N. Miller and R. A. de Callafon, “Identification of linear time-invariant systems via constrained step-based realization,” *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 1155 – 1160, 2012, 16th IFAC Symposium on System Identification.
- [13] H. Hjalmarsson, J. S. Welsh, and C. R. Rojas, “Identification of Box-Jenkins models using structured ARX models and nuclear norm relaxation,” *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 322 – 327, 2012, 16th IFAC Symposium on System Identification.
- [14] Z. Liu, A. Hansson, and L. Vandenberghe, “Nuclear norm system identification with missing inputs and outputs,” *Systems & Control Letters*, vol. 62, no. 8, pp. 605 – 612, 2013.
- [15] Z. Liu and L. Vandenberghe, “Interior-point method for nuclear norm approximation with application to system identification,” *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2010.
- [16] M. Zorzi and R. Sepulchre, “Factor analysis of moving average processes,” in *2015 European Control Conference (ECC)*, 2015, pp. 3579–3584.
- [17] —, “AR identification of latent-variable graphical models,” *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2327–2340, 2016.
- [18] K. Glover, “All optimal Hankel-norm approximations of linear multivariable systems and their L_∞ -error bounds,” *International Journal of Control*, vol. 39, no. 6, pp. 1115–1193, 1984.
- [19] S. Y. Kung, “A new identification and model reduction algorithm via singular value decomposition,” *12th Asilomar Conference on Circuits, Systems and Computers*, pp. 705–714, 1978.
- [20] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, “Structured low rank approximation,” *Linear Algebra and its Applications*, vol. 366, pp. 157 – 172, 2003.
- [21] A. Antoulas, *Approximation of Large-Scale Dynamical Systems*. SIAM, 2005.
- [22] V. Larsson and C. Olsson, “Convex low rank approximation,” *International Journal of Computer Vision*, vol. 120, no. 2, pp. 194–214, 2016.

- [23] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.
- [24] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, “Optimization with sparsity-inducing penalties,” *Foundations and Trends in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2012.
- [25] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, vol. 9, no. 6, p. 717, 2009.
- [26] A. Zare, M. R. Jovanović, and T. T. Georgiou, “Colour of turbulence,” *Journal of Fluid Mechanics*, vol. 812, pp. 636–680, 2017.
- [27] C. Grussler and A. Rantzer, “On optimal low-rank approximation of non-negative matrices,” in *54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 5278–5283.
- [28] P. L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*. Springer New York, 2011, pp. 185–212.
- [29] C. Grussler, “Rank reduction with convex constraints,” Ph.D. dissertation, Lund University, 2017.
- [30] C. Grussler and P. Giselsson, “Low-rank inducing norms with optimality interpretations,” arXiv:1612.03186v1, 2016.
- [31] —, “Local convergence of proximal splitting methods for rank constrained problems,” in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 702–708.
- [32] C. Grussler, “LRINorm,” <https://github.com/LowRankOpt/LRINorm>, 2018.
- [33] V. D. Blondel, E. D. Sontag, M. Vidyasagar, and J. C. Willems, *Open problems in mathematical systems and control theory*. Springer Science & Business Media, 2012.
- [34] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2012.
- [35] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I: Fundamentals*, ser. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1996.
- [36] —, *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, ser. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1993.
- [37] G. W. Stewart and J.-g. Sun, *Matrix Perturbation Theory*. Academic press, 1990.
- [38] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
- [39] A. Argyriou, R. Foygel, and N. Srebro, “Sparse prediction with the k-support norm,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1457–1465.
- [40] L. Jacob, G. Obozinski, and J.-P. Vert, “Group lasso with overlaps and graph lasso,” in *Proceedings of the 26th International Conference on Machine Learning*, L. Bottou and M. Littman, Eds. Montreal: Omnipress, June 2009, pp. 433–440.
- [41] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.
- [42] C. Berge, *Topological Spaces: Including a Treatment of Multi-Valued Functions, Vector Spaces, and Convexity*. Courier Corporation, 1963.
- [43] D. Peaucelle, D. Henrion, Y. Labit, and K. Taitz, “Users guide for SEDUMI INTERFACE 1.04,” 2002, LAAS-CNRS, Toulouse.
- [44] X. V. Doan and S. Vavasis, “Finding the largest low-rank clusters with Ky Fan 2-k-norm and ℓ_1 -Norm,” *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 274–312, 2016.