



**HAL**  
open science

## Hankel low-rank matrix completion: performance of the nuclear norm relaxation

Konstantin Usevich, Pierre Comon

► **To cite this version:**

Konstantin Usevich, Pierre Comon. Hankel low-rank matrix completion: performance of the nuclear norm relaxation. IEEE Journal of Selected Topics in Signal Processing, 2016, 10 (4), pp.637-646. <10.1109/JSTSP.2016.2535182>. <hal-01130631v3>

**HAL Id: hal-01130631**

**<https://hal.science/hal-01130631v3>**

Submitted on 17 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Hankel low-rank matrix completion: performance of the nuclear norm relaxation

Konstantin Usevich and Pierre Comon, *Fellow, IEEE*

**Abstract**—The completion of matrices with missing values under the rank constraint is a non-convex optimization problem. A popular convex relaxation is based on minimization of the nuclear norm (sum of singular values) of the matrix. For this relaxation, an important question is whether the two optimization problems lead to the same solution. This question was addressed in the literature mostly in the case of random positions of missing elements and random known elements. In this contribution, we analyze the case of structured matrices with a fixed pattern of missing values, namely, the case of Hankel matrix completion. We extend existing results on completion of rank-one real Hankel matrices to completion of rank- $r$  complex Hankel matrices.

**Index Terms**—Hankel; matrix completion; nuclear norm

## I. INTRODUCTION

THE problem of completing matrices with missing entries can be traced back to the works of Prony in 1795, and has been addressed since in various fields including: compressed sensing [1], [2], [3], system identification and control [4], [5], [6], graph theory [7], information theory [8], chemometrics [9], seismics [10], [11], estimation problems and sensor networks [12], to cite a few. It also appears as a subproblem in the computation of symmetric tensor Canonical Polyadic (CP) decompositions [13].

In the general case, the low-rank matrix completion problem is NP-hard (see, *e.g.* [14], [15]), although for some structured matrices the solution of the exact matrix completion problem can be found in polynomial time [16], [17], [18], [19], [20].

A popular approach in machine learning [14], [21] is to build a convex relaxation of the matrix completion problem by replacing the rank with the nuclear norm (*i.e.* the sum of singular values). There is a large body of research that showed that the nuclear norm relaxation can be successfully used both for exact and approximate matrix completion [2], [21], [3], [12], [1]. However, most of these results were proved in a random setting: the positions of missing entries are assumed to be drawn randomly; often the known entries are also assumed to be random. In addition, mostly unstructured matrices are considered, or the structured matrices when the linear constraints are distributed randomly.

In signal processing and system identification, the matrices under consideration are often structured [22] (the matrix structures are mostly Hankel/Toeplitz or alike), therefore the general results on the performance of the nuclear norm cannot be applied. Although the nuclear norm relaxation started to be actively used by practitioners, the question on its effectiveness

for structured problems was mostly unexplored [23]. Up to the authors' knowledge, there exists only one result [24], which treats the case of fixed (Hankel) structure and fixed pattern of missing values. This paper is a continuation of the work in that direction.

**Notation.** To ease the reading, vectors are denoted in boldface, *e.g.*  $\mathbf{p}$ , and matrices in bold capital, *e.g.*  $\mathbf{S}$ , or in calligraphic font when they are structured and parameterized, *e.g.*  $\mathcal{S}(\mathbf{p})$ .

### A. Problem statement

For a finite sequence  $\mathbf{h} = [h_0 \ \dots \ h_{2d}]^\top \in \mathbb{C}^{2d+1}$ , we denote the square Hankel matrix as

$$\mathcal{H}_d(\mathbf{h}) \stackrel{\text{def}}{=} [h_{k+l}]_{k,l=0}^{d,d} = \begin{bmatrix} h_0 & h_1 & \dots & h_d \\ h_1 & h_2 & \dots & h_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_d & h_{d+1} & \dots & h_{2d} \end{bmatrix}. \quad (1)$$

Occasionally,  $\mathbf{h}$  will be provided with a subscript indicating the index of its last entry; for instance in (1),  $\mathbf{h} = \mathbf{h}_{2d}$ . We are interested in the low-rank completion of (1), where the values  $h_0, \dots, h_d$  (highlighted by a gray background above) are known, and the other values ( $h_{d+1}, \dots, h_{2d}$ ) are to be completed.

Such a matrix completion problem appears, for example, in time series forecasting [25]. An extension of this problem (to block-Hankel matrices) appears in system and control theory: system realization [26] and system identification problems [27]. In this paper, we consider the exact Hankel matrix completion problem.

**Problem 1.** Given  $\mathbf{h}_d = [h_0 \ \dots \ h_d]^\top \in \mathbb{C}^{d+1}$ ,

$$\underset{h_{d+1}, \dots, h_{2d} \in \mathbb{C}}{\text{minimize}} \quad \text{rank } \mathcal{H}_d(\mathbf{h}), \quad (2)$$

where  $\mathbf{h} = [h_0 \ \dots \ h_{2d}]^\top$ .

The corresponding convex relaxation for Problem 1 is

**Problem 2.** Given  $\mathbf{h}_d = [h_0 \ \dots \ h_d]^\top \in \mathbb{C}^{d+1}$ ,

$$\underset{h_{d+1}, \dots, h_{2d} \in \mathbb{C}}{\text{minimize}} \quad \|\mathcal{H}_d(\mathbf{h})\|_*, \quad (3)$$

where  $\mathbf{h} = [h_0 \ \dots \ h_{2d}]^\top$ , and  $\|\cdot\|_*$  denotes the sum of singular values of a matrix (referred to as its nuclear norm).

Similarly to most of the papers on low-rank matrix recovery, we are interested in the following question:

**(Q)** When do the solutions of Problem 1 and Problem 2 coincide?

K. Usevich and P. Comon are with CNRS, GIPSA-Lab, Univ. Grenoble Alpes, F-38000 Grenoble, France. E-mail: `firstname.lastname@gipsa-lab.fr`.

Authors' version of the manuscript, accepted on February 7, 2016.

In practice, of course, the approximate case is more interesting (*i.e.* when we assume that we know the given values only approximately). We deliberately restrict ourselves to the exact matrix completion for two reasons. First, it is a prerequisite for studying the approximate matrix completion. Second, as it was shown in [12], a result on exact matrix completion implies that the nuclear norm can be applied in the approximate case if the noise is small enough.

### B. Previous results and contribution of the paper

Problem 1 is entirely solved [18], but the answer to our main question (Q) is known only for a class of rank-one cases.

**Theorem 1** ([24, Thm. 1]). *Let  $\mathbf{h}_d = [1 \ \lambda \ \dots \ \lambda^d]^\top$ , where  $\lambda \in (-1; 1)$ . Then the solution of*

$$\underset{h_{d+1}, \dots, h_{2d} \in \mathbb{R}}{\text{minimize}} \quad \|\mathcal{H}_d(\mathbf{h})\|_*, \quad (4)$$

*is unique and coincides with the minimal rank (rank-1) completion, which is given by  $p_k = \lambda^{d+k}$ .*

In this paper we extend the results of [24] to Hankel matrices with arbitrary rank. As in [24], we consider cases when the solution to Problem 1 is known, and establish conditions on  $\mathbf{h}_d$  when the solutions of Problem 1 and Problem 2 coincide.

Our results are more general (for example, applicable both in the real and complex cases), and do not depend on the results of [24]. However, we believe that our exposition may also help to understand the complicated logic of the proof in [24]. Moreover, this approach is also applicable to quasi-Hankel matrices [28].

In Section II we summarize the solution of Problem 1, and formally state the main results of the paper. In Section III we recall optimality conditions for convex optimization problems, and specialize them to nuclear norm minimization for complex structured matrices. Section IV contains the main lemmas and proofs of the paper. The lemmas imply that the optimality conditions are satisfied if the column space of completed matrices is not too far from a certain simply structured subspace. Finally, we illustrate the main results of the paper with numerical experiments in Section V.

## II. HANKEL MATRIX COMPLETION

### A. Solution of Hankel low-rank completion

Probably the first complete solution of Problem 1 is contained in [17, Ch. II], where the problem of low-rank matrix completion is called ‘‘singular extension’’ of Hankel matrices. In this section, we will use the theory from [18], and we will provide a summary of results in a simplified form.

For a finite sequence  $\mathbf{h}_d = [h_0 \ \dots \ h_d]^\top \in \mathbb{C}^{d+1}$ , we denote by  $\text{hrank}(\mathbf{h}_d)$  the smallest  $r$  such that there exists a non-zero vector  $\mathbf{q} = [q_0 \ q_1 \ \dots \ q_r]^\top \neq \mathbf{0}$  for which

$$q_0 h_k + \dots + q_r h_{k+r} = 0, \quad \forall k \in \{0, \dots, d-r\}. \quad (5)$$

The value  $\text{hrank}(\mathbf{h}_d)$  is equal to the maximal rank of the submatrices of the known triangular part of the Hankel matrix (1) (or, equivalently to the maximal rank of a Hankel matrix that can be constructed only from the elements of the vector

$\mathbf{h}_d$ ). In [18, Def. 5.3, page 81], the number  $\text{hrank}(\mathbf{h}_d)$  is called *the first characteristic degree* of  $\mathbf{h}_d$ . We will also call the vector  $\mathbf{q}$  as *the characteristic vector*<sup>1</sup> of  $\mathbf{h}_d$ . Next, we recall the following basic properties of  $\mathbf{h}_d$  and  $\mathbf{q}$ .

- For any  $\mathbf{h}_d \in \mathbb{C}^{d+1}$ ,  $\text{hrank}(\mathbf{h}_d) \leq \frac{d+2}{2}$  ([18, Prop. 5.4]).
- If  $\text{hrank}(\mathbf{h}_d) < \frac{d+2}{2}$ , then the characteristic vector  $\mathbf{q}$  is unique up to scaling by a constant (see [18, page 84]).

In what follows, we assume that  $q_r \neq 0$  (this is a generic case). The treatment of the case  $q_r = 0$  can be found in [18, page 84]). Then an explicit representation of  $\mathbf{h}_d$  is determined by the *characteristic polynomial*

$$q(z) \stackrel{\text{def}}{=} \sum_{j=0}^r q_j z^j = q_r z^{\nu_0} \prod_{k=1}^s (z - \lambda_k)^{\nu_k}, \quad (6)$$

where  $\lambda_k \in \mathbb{C} \setminus \{0\}$  are distinct,  $\nu_0 \geq 0$ ,  $\nu_k \geq 1$  are the multiplicities of the roots, and  $r = \nu_0 + \nu_1 + \dots + \nu_s$ .

**Proposition 2** (A special case of [18, Thm. 8.1, p.124]). *If  $q_r \neq 0$ ,  $\nu_0 = 0$  and all the roots of  $q(z)$  are simple ( $\nu_k = 1$ ,  $s = r$ ), then the sequence  $\mathbf{h}_d$  admits a representation*

$$h_k = \sum_{j=1}^r c_j \lambda_j^k, \quad (7)$$

where  $c_j \in \mathbb{C}$ .

**Example 1.** For  $\mathbf{h}_d = [1 \ \lambda \ \dots \ \lambda^d]^\top$  we have  $\text{hrank}(\mathbf{h}_d) = 1$  and the characteristic vector can be chosen as  $\mathbf{q} = [-\lambda \ 1]^\top$ .

The representation (7) is called a *canonical representation* of  $\mathbf{h}_d$  (or the unique canonical representation in case  $\text{hrank}(\mathbf{h}_d) < \frac{d+2}{2}$ ). In the general case (6), there are multiple roots and the canonical representation has a more complicated form, as stated by the proposition below.

**Proposition 3** (A special case of [18, Thm. 8.1, p.124]). *Let  $q_r \neq 0$  and  $\lambda_k, \nu_k$  are as in (6). Then  $\mathbf{h}_d$  has the form*

$$h_k = \sum_{j=1}^s c_j(k) \lambda_j^k + \sum_{l=0}^{\nu_0-1} c_{0,l} \delta(k, l), \quad (8)$$

where  $c_j(k)$  are polynomials of degree  $(\nu_j - 1)$ ,  $c_{0,l} \in \mathbb{C}$ , and  $\delta(k, l)$  is the Kronecker delta.

**Remark 4.** In the statement of Proposition 3, if  $\nu_0 = 0$  then the second term in (8) is absent.

**Example 2.** In the extreme case  $q(z) = q_r z^r$  (*i.e.*  $r = \nu_0$ ), the canonical representation (8) becomes

$$\mathbf{h}_d = [c_{0,0} \ \dots \ c_{0,r-1} \ 0 \ \dots \ 0]^\top, \quad (9)$$

Apart from the explicit form of the sequence  $\mathbf{h}_d$ , the characteristic polynomial gives a solution to the rank minimization problem (2) (for  $m = 1$ ).

**Proposition 5** (A corollary of [18, Thm. 5.14, p.99]). *Let  $\mathbf{q} \in \mathbb{C}^{r+1}$  be a characteristic vector of  $\mathbf{h}_d$  with  $q_r \neq 0$ ,  $r = \text{hrank}(\mathbf{h}_d)$ . Then it holds that*

<sup>1</sup>In [18],  $\mathbf{q}$  does not have a specific name and is denoted by  $p$ .

- (i) The rank of the minimal rank completion in (2) is  $r$ .  
 (ii) A minimum rank completion  $h_{d+1}, \dots, h_{2d}$  is given by the recursive continuation, for  $k > d - r$ :

$$h_{k+r} = -\frac{1}{q_r}(q_0 h_k + \dots + q_{r-1} h_{k+r-1}) \quad (10)$$

In addition, the values of  $h_{k+r}$  for  $k > d - r$  can be obtained by using the corresponding formula of canonical representation, namely (7) or (8).

- (iii) If  $r < \frac{d+2}{2}$ , then the minimum rank completion is unique.

### B. Main results

First, we refine Theorem 1.

**Theorem 6.** Let  $\mathbf{h}_d = [1 \ \lambda \ \dots \ \lambda^d]^\top$ , with  $\lambda \in \mathbb{C}$ .

(i) If  $|\lambda| < 1$ , the solution of (3) is unique and coincides with the minimal rank (rank-1) completion, which is given by  $h_{d+k} = \lambda^{d+k}$ .

(ii) Let  $|\lambda| \geq 1$ , with  $\lambda = \rho e^{i\theta}$  and  $\rho \in \mathbb{R}$  (which implies  $h_k = \rho^k e^{i\theta k}$  for  $k \in \{0, \dots, d\}$ ). Then a solution of (3) is given by  $h_{d+k} = \rho^{d-k} e^{i\theta(d+k)} = \lambda^d / (\bar{\lambda})^k$ .

(iii) If  $|\lambda| > 1$  the canonical completion  $h_{d+k} = \lambda^{d+k}$  is not a minimizer of (3).

(iv) If  $\lambda \in \mathbb{R}$ , then the conclusions of (i)–(iii) are valid for the real nuclear norm minimization problem (4).

Next, we prove a result that generalizes Theorem 1 to rank- $r$  Hankel matrices.

**Theorem 7.** For any  $d$ , and  $r < \frac{d+2}{2}$  there exists  $\rho_{d,r} > 0$  such that for all  $\mathbf{h}_d \in \mathbb{C}^{d+1}$  with  $\text{hrank}(\mathbf{h}_d) = r$  and  $|\lambda_k| \leq \rho_{d,r}$  in (6), it holds that:

- (i) the solution of (3) is unique and coincides with (10);  
 (ii) if  $\mathbf{h}_d \in \mathbb{R}^{d+1}$ , the solution of (4) is unique and coincides with (10).

The proofs of both theorems can be found in Section IV.

**Remark 8.** Theorem 6 proves a special case of Theorem 7 for  $r = 1$  and guarantees that  $\rho_{d,1}$  can be chosen slightly less than 1.

## III. OPTIMALITY CONDITIONS OF NUCLEAR NORM MINIMIZATION

In this section, we recall some definitions and basic results from the field of convex optimization [29, Chapter D], and specialize them to our problem.

### A. Optimality conditions in convex optimization

For a convex (possibly non-differentiable) function  $f : \mathbb{R}^N \rightarrow \mathbb{R}$ , the subdifferential [29, Def. 1.2.1] of  $f$  is defined as the set  $\partial f(\mathbf{x}) \subset \mathbb{R}^N$

$$\partial f(\mathbf{x}) \stackrel{\text{def}}{=} \{\mathbf{d} : f(\mathbf{y}) - f(\mathbf{x}) \geq \langle \mathbf{d}, \mathbf{y} - \mathbf{x} \rangle \quad \forall \mathbf{y} \in \mathbb{R}^N\}.$$

In particular, if  $f$  is differentiable at a point  $\mathbf{x}$ , then the sub-gradient has only one element: the usual gradient, i.e.  $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$ .

For the unconstrained convex optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}), \quad (11)$$

we recall the first-order optimality condition, which is necessary and sufficient in this case.

**Lemma 9** (First-order optimality, [29, Thm. 2.2.1]). A point  $\mathbf{x}$  is a minimum point of (11) if and only if

$$\mathbf{0} \in \partial f(\mathbf{x}). \quad (12)$$

In particular, for differentiable functions, the condition of Lemma 9 reduces to  $\nabla f(\mathbf{x}) = \mathbf{0}$ . Next, we use a simple sufficient condition for uniqueness of the minimizer.

**Lemma 10** (Sufficient condition of uniqueness). A point  $\mathbf{x}$  is the unique minimizer of (11) if

$$\mathbf{0} \in \text{int}(\partial f(\mathbf{x})), \quad (13)$$

where  $\text{int}(\cdot)$  denotes the interior of a set.

*Proof.* In this case for any  $\mathbf{y} \neq \mathbf{x}$  there exists  $\delta > 0$  such that  $\delta \cdot (\mathbf{y} - \mathbf{x}) \in \partial f(\mathbf{x})$ . By definition of  $\partial f(\mathbf{x})$ , we have that  $f(\mathbf{y}) - f(\mathbf{x}) \geq \delta \|\mathbf{y} - \mathbf{x}\|_2^2 > 0$ .  $\square$

### B. Subdifferential of the nuclear norm

First, we recall the form of the subdifferential of the nuclear norm of a matrix. Let  $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$  be a matrix of rank  $r$ , and let  $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$  be an SVD of  $\mathbf{X}$ , where  $\mathbf{U} \in \mathbb{R}^{n_1 \times r}$ ,  $\mathbf{V} \in \mathbb{R}^{n_2 \times r}$  and  $\mathbf{\Sigma} \in \mathbb{R}^{r \times r}$  is a diagonal matrix of nonzero singular values. Next, let  $\mathbf{U}_\perp \in \mathbb{R}^{n_1 \times (n_1 - r)}$ ,  $\mathbf{V}_\perp \in \mathbb{R}^{n_2 \times (n_2 - r)}$  be some orthonormal bases of the left and right nullspaces of  $\mathbf{X}$ , respectively. Then, according to [30, p. 41], the subdifferential of the nuclear norm at  $\mathbf{X}$  is equal to  $\partial \|\mathbf{X}\|_* =$

$$\{\mathbf{B} + \mathbf{U}_\perp \mathbf{W} \mathbf{V}_\perp^\top : \mathbf{W} \in \mathbb{R}^{(n_1 - r) \times (n_2 - r)}, \|\mathbf{W}\|_2 \leq 1\},$$

where  $\mathbf{B}$  is defined as

$$\mathbf{B} \stackrel{\text{def}}{=} \mathbf{U} \mathbf{V}^\top, \quad (14)$$

and  $\|\cdot\|_2$  is the spectral norm (largest singular value).

**Remark 11.** When the SVD of  $\mathbf{X}$  is not unique, then the matrix  $\mathbf{B}$  from (14) is still defined uniquely. In fact, if  $\mathbf{X}$  has multiple singular values, its matrix of left (resp. right) singular vectors is of the form  $\mathbf{U}\mathbf{\Theta}$  (resp.  $\mathbf{V}\mathbf{\Theta}$ ), where  $\mathbf{\Theta} \in \mathbb{R}^{r \times r}$  is an orthogonal matrix which commutes with  $\mathbf{\Sigma}$ .

**Remark 12.** Let  $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$  be an SVD of  $\mathbf{X}$ . When  $\mathbf{X}$  is real symmetric, we do not necessarily have that  $\mathbf{U} = \mathbf{V}$ . In fact, it is only true when  $\mathbf{X}$  is either positive or negative semidefinite. That is why  $\mathbf{B} \neq \mathbf{U}\mathbf{U}^\top$  even when  $\mathbf{X}$  is real symmetric.

### C. Real nuclear norm minimization

In this section, we consider a real nuclear norm minimization problem

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p} \in \mathbb{R}^N} f(\mathbf{p}), \quad f(\mathbf{p}) \stackrel{\text{def}}{=} \|\mathcal{S}(\mathbf{p})\|_*, \quad (15)$$

where  $\mathcal{S}$  is an affine map<sup>2</sup>  $\mathbb{R}^N \rightarrow \mathbb{R}^{n_1 \times n_2}$  defined as

$$\mathcal{S}(\mathbf{p}) = \mathbf{S}_0 + \sum_{k=1}^N p_k \mathbf{S}_k, \quad (16)$$

<sup>2</sup>In [22] and other literature, the map  $\mathcal{S}$  is often called *matrix structure*.

and where  $\mathbf{S}_k$ ,  $k \in \{0, \dots, N\}$  are known linearly independent real  $n_1 \times n_2$  matrices, and  $\mathbf{p} = [p_1, \dots, p_N]^T$ .

Next, define  $\mathbf{Q}_1 = \mathbf{U}_\perp \mathbf{U}_\perp^T$  and  $\mathbf{Q}_2 = \mathbf{V}_\perp \mathbf{V}_\perp^T$  (the projectors on the column and row space of  $\mathbf{X}$ ). Then the subdifferential can be rewritten as

$$\partial \|\mathbf{X}\|_* = \{\mathbf{B} + \mathbf{Q}_1 \mathbf{M} \mathbf{Q}_2 : \|\mathbf{M}\|_2 \leq 1, \mathbf{M} \in \mathbb{R}^{n_1 \times n_2}\}, \quad (17)$$

where  $\mathbf{B}$  is defined as in (14).

Finally, by an analogue of the chain rule [29, Thm. 4.2.1], we immediately have that

$$\begin{aligned} \partial f(\mathbf{p}) &= \left\{ [\langle \mathbf{S}_1, \mathbf{H} \rangle_F \ \cdots \ \langle \mathbf{S}_N, \mathbf{H} \rangle_F]^T : \right. \\ &\quad \left. \mathbf{H} \in \partial \|\mathbf{X}\|_*, \mathbf{X} = \mathcal{S}(\mathbf{p}) \right\}, \quad (18) \end{aligned}$$

where  $\langle \cdot, \cdot \rangle_F$  denotes the Frobenius inner product.

#### D. Complex nuclear norm minimization: symmetric matrices

Now consider the map  $\mathcal{S}(\mathbf{p})$  given in (16) but with  $\mathbf{S}_0 \in \mathbb{C}^{n \times n}$ ,  $\mathbf{S}_k \in \mathbb{R}^{n \times n}$  for  $k \geq 1$  (all  $\mathbf{S}_k$  are symmetric) and with a complex parameter  $\mathbf{p} \in \mathbb{C}^N$ , i.e. (16) defines a complex-valued map  $\mathcal{S} : \mathbb{C}^N \rightarrow \mathbb{C}^{n \times n}$ . We also consider the corresponding nuclear norm minimization problem

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p} \in \mathbb{C}^N} \|\mathcal{S}(\mathbf{p})\|_*. \quad (19)$$

In what follows we derive the optimality conditions for (19) from Lemmas 9 and 10.

Let  $\mathcal{S}(\mathbf{p}) = \mathbf{U} \Sigma \mathbf{V}^H$  be an SVD of the symmetric rank- $r$  matrix  $\mathcal{S}(\mathbf{p})$  (with  $\mathbf{U}, \mathbf{V} \in \mathbb{C}^{n \times r}$ ) and define

$$\mathbf{B} \stackrel{\text{def}}{=} \mathbf{U} \mathbf{V}^H, \quad \mathbf{P} \stackrel{\text{def}}{=} \mathbf{U} \mathbf{U}^H = \mathbf{B} \mathbf{B}^H, \quad (20)$$

where  $\mathbf{B}$  is analogous to the matrix  $\mathbf{B}$  in (14), and  $\mathbf{P}$  is the orthogonal projector on the column space of  $\mathcal{S}(\mathbf{p})$ .

**Remark 13.** For a complex symmetric matrix  $\mathcal{S}(\mathbf{p})$ , the matrix  $\mathbf{B}$  defined in (20) is also complex symmetric. In addition,  $\mathbf{P}^T = \mathbf{V} \mathbf{V}^H$ . As in Remark 12,  $\mathbf{B}$  is not necessarily equal to  $\mathbf{P}$ .

Next, we define the matrix

$$\mathbf{S} \stackrel{\text{def}}{=} [\text{vec}(\mathbf{S}_1) \ \cdots \ \text{vec}(\mathbf{S}_N)], \quad (21)$$

and for a matrix  $\mathbf{P} \in \mathbb{C}^{n \times n}$  we define

$$\mathcal{A}(\mathbf{P}) \stackrel{\text{def}}{=} \mathbf{S}^T ((\mathbf{I} - \mathbf{P}) \boxtimes (\mathbf{I} - \mathbf{P})) \in \mathbb{C}^{N \times n^2}, \quad (22)$$

where  $\boxtimes$  denotes the Kronecker product.

The following proposition shows that we can write the optimality conditions (as in Lemmas 9 and 10) for the problem (19) in a convenient form.

**Proposition 14** (Complex optimality conditions). *Let  $\mathbf{P}, \mathbf{B}$  be as in (20). Then it holds that*

(i) *The point  $\mathbf{p}$  is a minimizer of (19) if and only if there exists  $\mathbf{M} \in \mathbb{C}^{n \times n}$  with  $\|\mathbf{M}\|_2 \leq 1$  that satisfies*

$$\mathcal{A}(\mathbf{P}) \text{vec}(\mathbf{M}) = -\mathbf{S}^T \text{vec}(\mathbf{B}). \quad (23)$$

(ii) *If, in addition, the matrix  $\mathbf{M}$  satisfies  $\|\mathbf{M}\|_2 < 1$  and  $\mathcal{A}(\mathbf{P})$  is full row rank, then the point  $\mathbf{p}$  is the unique minimizer of (19).*

*Proof.* See Appendix A for a proof.  $\square$

**Remark 15.** *If  $\mathbf{p} \in \mathbb{R}^n$  and  $\mathbf{S}_0 \in \mathbb{R}^{n \times n}$ , for  $f(\mathbf{p})$  defined in (15), the subgradient (18) is equal to*

$$\partial f(\mathbf{p}) = \{\mathbf{S}^T \text{vec}(\mathbf{B}) + \mathcal{A}(\mathbf{P}) \text{vec}(\mathbf{M}) : \|\mathbf{M}\|_2 \leq 1\}, \quad (24)$$

with  $\mathbf{M} \in \mathbb{R}^{n \times n}$  and  $\mathbf{I} - \mathbf{P} = \mathbf{Q}_1 = \mathbf{Q}_2$ . Hence, Proposition 14 gives optimality conditions also for the problem (15) (the only difference is that  $\mathbf{M}$  should be real).

**Remark 16.** *The condition  $\text{rank } \mathcal{A}(\mathbf{P}) = N$  is equivalent to  $\mathbf{Q} \mathbf{H} \mathbf{Q}^T \neq \mathbf{0}$  for  $\mathbf{Q} = \mathbf{I} - \mathbf{P}$  and any  $\mathbf{H}$  of the form*

$$\mathbf{H} = \sum_{k=1}^N \Delta p_k \mathbf{S}_k, \quad \Delta p \in \mathbb{C}^N \setminus \{\mathbf{0}\},$$

which corresponds to [24, Proposition 2]. Next, the condition (23) is equivalent to

$$\langle \mathbf{S}_k, \mathbf{B} + \mathbf{Q} \mathbf{M} \mathbf{Q}^T \rangle_F = 0, \quad \forall k \in \{1, \dots, N\}. \quad (25)$$

**Remark 17.** *Remarks 15 and 16 are related to the optimality conditions used in [24]. Indeed, the condition  $\mathbf{Q} \mathbf{H} \mathbf{Q}^T \neq \mathbf{0}$  appears in [24, Proposition 2]. The condition (25) corresponds to the condition in [24, Lemma 2]. The authors of [24], however, do not make explicit connection with general optimality conditions for convex optimization problems, presented in Section III-A.*

## IV. MAIN LEMMAS AND RESULTS

In this section, we derive the proofs of Theorems 6 and 7. The idea is to start from simple cases where optimality conditions can be easily verified, and then show (in Section IV-C) that under a sufficiently small perturbation, these optimality conditions still hold.

### A. Hankel matrix completion: basis matrices

First, we put the nuclear norm minimization problem (2) in the form (19). In order to do this, we explicitly write down the matrices  $\mathbf{S}_k$  in (16). The constant part has the form

$$\mathbf{S}_0 = \begin{bmatrix} h_0 & h_1 & \cdots & h_d \\ h_1 & h_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_d & 0 & \cdots & 0 \end{bmatrix}.$$

The matrices  $\mathbf{S}_k$ ,  $k \in \{1, \dots, N\}$ , with  $N = d$ ,  $n = d+1$  are defined as

$$\begin{aligned} \mathbf{S}_1 &= \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & 0 & \vdots \\ 0 & 1 & 0 & \cdots & 0 \end{bmatrix}, \dots \\ \mathbf{S}_{N-1} &= \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \mathbf{S}_N = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}, \quad (26) \end{aligned}$$

such  $\mathcal{S}(\mathbf{p}) = \mathcal{H}_d(\mathbf{h})$  for  $\mathbf{p} = [h_{d+1}, \dots, h_{2d}]^T$ .

### B. Simple matrices

Now we show that optimality conditions hold true for simple matrices.

**Lemma 18.** *Let  $\mathbf{P}_0 \in \mathbb{C}^{n \times n}$  be a matrix of the form*

$$\mathbf{P}_0 \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{P}'_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{P}'_0 \in \mathbb{C}^{r \times r}, \quad (27)$$

where  $r \leq \frac{d+2}{2}$ . Then it holds that

(i) For any  $k \in \{1, \dots, N\}$ ,  $\mathbf{P}_0^\top \mathbf{S}_k \mathbf{P}_0 = \mathbf{0}$ .

(ii) If, in addition,  $r < \frac{d+2}{2}$ , then  $\text{rank } \mathcal{A}(\mathbf{P}_0) = N$ .

*Proof.* (i) Since  $r \leq \frac{d+2}{2}$ , then from (26) we have that for any  $k \in \{1, \dots, N\}$

$$\mathbf{S}_k = \left[ \begin{array}{c|c} \mathbf{0}_{r \times r} & * \\ \hline * & * \end{array} \right]. \quad (28)$$

Therefore,  $\mathbf{P}_0^\top \mathbf{S}_k \mathbf{P}_0 = \mathbf{0}$ .

(ii) Now, as in Remark 16, we take  $\mathbf{H} = \sum_{k=1}^N \Delta p_k \mathbf{S}_k \neq \mathbf{0}$ , define the matrix  $\mathbf{Q}_0$

$$\mathbf{Q}_0 \stackrel{\text{def}}{=} \mathbf{I}_n - \mathbf{P}_0 = \begin{bmatrix} \mathbf{I}_r - \mathbf{P}'_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-r} \end{bmatrix},$$

and we need to prove that  $\mathbf{Q}_0 \mathbf{H} \mathbf{Q}_0^\top \neq \mathbf{0}$ .

Since  $r < \frac{d+2}{2}$ , we have that  $\lfloor \frac{d}{2} \rfloor + 1 \leq n - r$ ,  $\lceil \frac{d}{2} \rceil \leq n - r$ , and therefore

$$\mathbf{Q}_0 \mathbf{H} \mathbf{Q}_0^\top = \begin{bmatrix} * & * \\ * & \mathbf{F} \end{bmatrix},$$

where

$$\mathbf{F} = \begin{bmatrix} \Delta p_1 & \Delta p_2 & \cdots & \Delta p_{\lceil \frac{d}{2} \rceil} \\ \Delta p_2 & \ddots & \ddots & \Delta p_{\lceil \frac{d}{2} \rceil + 1} \\ \vdots & \ddots & \ddots & \vdots \\ \Delta p_{\lfloor \frac{d}{2} \rfloor + 1} & \ddots & \ddots & \Delta p_d \end{bmatrix}.$$

Hence, for all  $\Delta p \neq \mathbf{0}$  we have that  $\mathbf{Q}_0 \mathbf{H} \mathbf{Q}_0^\top \neq \mathbf{0}$ .  $\square$

From Lemma 18, we can deduce the following corollary, which answers question (Q) for simple cases.

**Corollary 19.** *In the completion problem (2), let the vector  $\mathbf{h}_d$  be of the form (9) with  $r \leq \frac{d+2}{2}$ . Then we have the following.*

(i) A solution of the rank minimization problem is given by  $h_k = 0$  for  $k > d$ , and coincides with the solution of (3).

(ii) If, in addition,  $r < \frac{d+2}{2}$ , the solution of (3) is unique.

*Proof.* (i) Let  $\mathbf{P}_0$  be as in (27), and  $\mathbf{P}'_0 = \mathbf{I}_r$ . Then, the matrix  $\mathbf{B}$  defined in (20) (for  $\mathcal{S}(\mathbf{p}) = \mathcal{H}_d(\mathbf{h})$ ) has the form  $\mathbf{B} = \mathbf{P}_0 \mathbf{A} \mathbf{P}_0^\top$ . By Lemma 18, we have that  $\langle \mathbf{S}_k, \mathbf{B} \rangle_F = \langle \mathbf{P}_0^\top \mathbf{S}_k \mathbf{P}_0, \mathbf{A} \rangle_F = 0$ . The rest follows from Remark 16.

(ii) Follows from Remark 16.  $\square$

### C. Perturbations of simple projectors

Now we show that the conditions of Proposition 14 hold true for perturbations of simple matrices. First, we need a simple inequality on distance between projectors.

**Lemma 20.** *Let  $\mathbf{U}, \mathbf{V} \in \mathbb{C}^{n \times r}$  be such that  $\mathbf{U}^\top \mathbf{U} = \mathbf{V}^\top \mathbf{V} = \mathbf{I}_r$ ,  $\mathbf{B}$  and  $\mathbf{P}$  be defined in (20), and assume that  $\mathbf{B} = \mathbf{B}^\top$ . Let  $\mathbf{P}_0 \in \mathbb{C}^{n \times n}$  be an orthogonal projector with  $\text{rank}(\mathbf{P}_0) = r$ . Then we have that*

$$\|\mathbf{B} - \mathbf{P}_0 \mathbf{B} \mathbf{P}_0^\top\|_F^2 \leq 2\|(\mathbf{I} - \mathbf{P}_0)\mathbf{U}\|_F^2 = \|\mathbf{P} - \mathbf{P}_0\|_F^2.$$

*Proof.* Since  $\mathbf{B} - \mathbf{P}_0 \mathbf{B} \mathbf{P}_0^\top = \mathbf{B} - \mathbf{P}_0 \mathbf{B} + \mathbf{P}_0 \mathbf{B} - \mathbf{P}_0 \mathbf{B} \mathbf{P}_0^\top$ ,

$$\begin{aligned} \|\mathbf{B} - \mathbf{P}_0 \mathbf{B} \mathbf{P}_0^\top\|_F^2 &\leq \|(\mathbf{I} - \mathbf{P}_0)\mathbf{B}\|_F^2 + \|\mathbf{P}_0 \mathbf{B}(\mathbf{I} - \mathbf{P}_0^\top)\|_F^2 \\ &\leq 2\|(\mathbf{I} - \mathbf{P}_0)\mathbf{B}\|_F^2 = 2\|(\mathbf{I} - \mathbf{P}_0)\mathbf{U}\|_F^2 \end{aligned}$$

Finally,

$$\begin{aligned} \|\mathbf{P} - \mathbf{P}_0\|_F^2 &= \text{trace}\{\mathbf{P}^2 + \mathbf{P}_0^2 - \mathbf{P} \mathbf{P}_0 - \mathbf{P}_0 \mathbf{P}\} \\ &= 2 \text{trace}\{(\mathbf{I} - \mathbf{P}_0)\mathbf{P}\} = 2\|(\mathbf{I} - \mathbf{P}_0)\mathbf{U}\|_F^2, \end{aligned} \quad (29)$$

where the second equality above follows from the identities  $\text{trace}\{\mathbf{P}_0\} = \text{trace}\{\mathbf{P}\} = r$  and  $\text{trace}\{\mathbf{P} \mathbf{P}_0\} = \text{trace}\{\mathbf{P}_0 \mathbf{P}\}$ , and the last from idempotence of projectors.  $\square$

Next, we prove a perturbation lemma that uses a bound on the distance between projectors.

**Lemma 21.** *Let  $\mathbf{S}$  and  $\mathcal{A}(\mathbf{P})$  be defined in (21) and (22). Let  $\mathbf{P}_0 \in \mathbb{C}^{n \times n}$  be a rank- $r$  projector matrix such that  $\text{rank } \mathcal{A}(\mathbf{P}_0) = N$ ,  $\mathbf{P}_0^\top \mathbf{S}_k \mathbf{P}_0 = \mathbf{0}$  for all  $k \in \{1, \dots, N\}$ .*

*Then there exists a constant  $\varepsilon = \varepsilon_{d,r}$  such that for any  $\mathbf{B}, \mathbf{P}$  as in Lemma 20, which satisfy*

$$\|\mathbf{P} - \mathbf{P}_0\|_F \leq \varepsilon_{d,r},$$

*it holds that*

(i) *The matrix  $\mathcal{A}(\mathbf{P})$  is full row rank.*

(ii) *There exists a matrix  $\mathbf{M} \in \mathbb{C}^{n \times n}$ ,  $\|\mathbf{M}\|_2 < 1$ , such that (23) is satisfied.*

*Proof.* (i) Since  $\mathcal{A}(\mathbf{P})$  depends polynomially on the entries of  $\mathbf{P}$ , and  $\text{rank } \mathcal{A}(\mathbf{P}_0) = N$ , by lower semicontinuity of matrix rank, there exists a neighborhood

$$\mathcal{V}_{\varepsilon_1} \stackrel{\text{def}}{=} \{\mathbf{P} \in \mathbb{C}^{n \times n} : \|\mathbf{P} - \mathbf{P}_0\| \leq \varepsilon_1\},$$

such that  $\text{rank } \mathcal{A}(\mathbf{P}) = N$  for any  $\mathbf{P} \in \mathcal{V}_{\varepsilon_1}$ .

(ii) Take  $\varepsilon_1$  as in (i), and denote

$$\delta_1 \stackrel{\text{def}}{=} \inf_{\mathbf{P} \in \mathcal{V}_{\varepsilon_1}} \sigma_{\min}(\mathcal{A}(\mathbf{P})).$$

Since  $\mathcal{V}_{\varepsilon_1}$  is compact, we have  $\delta_1 > 0$ .

In fact, we can constructively find such  $\varepsilon_1$  and  $\delta_1$ . From the properties of singular values we have that

$$\begin{aligned} \sigma_{\min}^2(\mathcal{A}(\mathbf{P})) &\geq \sigma_{\min}^2(\mathcal{A}(\mathbf{P}_0)) - \|\mathcal{A}(\mathbf{P}) - \mathcal{A}(\mathbf{P}_0)\|_F^2 \\ &\geq 1 - \|\mathbf{S}\|_2^2 \|(\mathbf{I} - \mathbf{P}) \boxtimes (\mathbf{I} - \mathbf{P}) - (\mathbf{I} - \mathbf{P}_0) \boxtimes (\mathbf{I} - \mathbf{P}_0)\|_F^2 \\ &= 1 - 2\|\mathbf{S}\|_2^2 ((n-r)^2 - (\text{trace}\{(\mathbf{I} - \mathbf{P}_0)(\mathbf{I} - \mathbf{P})\})^2), \end{aligned}$$

where the last equality is derived similarly to (29) and using the identity  $\text{trace}\{A \boxtimes A\} = (\text{trace}\{A\})^2$ . Next, from (29),

$$\text{trace}\{(\mathbf{I} - \mathbf{P}_0)(\mathbf{I} - \mathbf{P})\} = (n-r) - \frac{\|\mathbf{P} - \mathbf{P}_0\|_F^2}{2}.$$

Therefore,

$$\sigma_{\min}^2(\mathcal{A}(\mathbf{P})) \geq 1 - 2(n-r)\|\mathbf{S}\|_2^2\|\mathbf{P} - \mathbf{P}_0\|_F^2,$$

and

$$\delta_1 \geq 1 - \sqrt{2(n-r)}\|\mathbf{S}\|_{2\varepsilon_1}. \quad (30)$$

Next, in the  $\varepsilon_1$ -neighborhood, a solution of (23) always exists because  $\mathcal{A}(\mathbf{P})$  is full row rank. Consider the minimum Frobenius norm solution  $\mathbf{M}_* = \mathbf{M}_*(\mathbf{P})$  of (23), given by

$$\begin{aligned} \text{vec}(\mathbf{M}_*(\mathbf{P})) &= -\mathcal{A}(\mathbf{P})^\dagger \mathbf{S}^\top \text{vec}(\mathbf{B}) \\ &= -\mathcal{A}(\mathbf{P})^\dagger \mathbf{S}^\top \text{vec}(\mathbf{B} - \mathbf{P}_0 \mathbf{B} \mathbf{P}_0^\top), \end{aligned}$$

where the matrix  $\mathcal{A}(\mathbf{P})^\dagger$  is the pseudoinverse of  $\mathcal{A}(\mathbf{P})$ , and the last equality holds since  $\mathbf{P}_0^\top \mathbf{S}_k \mathbf{P}_0 = 0$  for all  $k \in \{1, \dots, N\}$ . Then, by Lemma 20,

$$\begin{aligned} \|\mathbf{M}_*\|_2 &\leq \|\mathbf{M}_*\|_F = \|\text{vec}(\mathbf{M}_*)\|_2 \leq \\ &\|\mathcal{A}(\mathbf{P})^\dagger\|_2 \|\mathbf{S}^\top\|_2 \|\mathbf{B} - \mathbf{P}_0 \mathbf{B} \mathbf{P}_0^\top\|_F \leq \frac{\|\mathbf{S}\|_2}{\delta_1} \|\mathbf{P} - \mathbf{P}_0\|_F. \end{aligned} \quad (31)$$

Finally, define  $\varepsilon_{d,r} \stackrel{\text{def}}{=} \min(\varepsilon_1, \frac{\alpha \delta_1}{\|\mathbf{S}\|_2})$ , where  $0 < \alpha < 1$ . For such an  $\varepsilon_{d,r}$ , we have  $\|\mathbf{M}_*\|_* < 1$  if  $\|\mathbf{P} - \mathbf{P}_0\|_F \leq \varepsilon_{d,r}$ , which completes the proof.  $\square$

#### D. Proof of Theorem 6

Here we provide a complete picture for rank-one Hankel matrix completion. First, we need a technical lemma.

**Lemma 22.** For any complex matrix  $\mathbf{Y} \in \mathbb{C}^{n \times n}$ ,

$$\|\mathbf{Y}\|_* \geq |\text{trace}\{\mathbf{Y}\}|;$$

*Proof.* Consider an SVD of  $\mathbf{Y}$

$$\mathbf{Y} = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^H.$$

where  $\mathbf{u}_k, \mathbf{v}_k \in \mathbb{C}^n$ , and  $\|\mathbf{u}_k\|_2 = \|\mathbf{v}_k\|_2 = 1$ . Then

$$\begin{aligned} |\text{trace}\{\mathbf{Y}\}| &\leq \sum_{k=1}^r |\text{trace}\{\sigma_k \mathbf{u}_k \mathbf{v}_k^H\}| \\ &= \sum_{k=1}^r \sigma_k |\mathbf{v}_k^H \mathbf{u}_k| \leq \sum_{k=1}^r \sigma_k = \|\mathbf{Y}\|_*, \end{aligned}$$

which completes the proof.  $\square$

*Proof of Theorem 6.* As in Example 1, we denote by  $\mathbf{h}_d = [1 \ \lambda \ \dots \ \lambda^d]^\top$ . We are looking for minimizers of the nuclear norm of

$$\mathcal{S}(\mathbf{p}) = \begin{bmatrix} 1 & \lambda & \dots & \lambda^d \\ \lambda & \lambda^2 & \dots & p_1 \\ \vdots & \dots & \dots & \vdots \\ \lambda^d & p_1 & \dots & p_d \end{bmatrix}, \quad (32)$$

where  $n = d+1$  is the number of rows and columns of  $\mathcal{S}(\mathbf{p})$ .

We prove (i), (ii) and (iii) for real nonnegative  $\lambda$ , and proceed to the complex case (in part (v) of the proof).

(i) We need to show that for  $\lambda \in \mathbb{R}$ ,  $0 \leq \lambda < 1$  the canonical rank-one completion

$$\mathbf{p}_\lambda = [\lambda^{d+1} \ \dots \ \lambda^{2d}]^\top, \quad (33)$$

is the unique solution of (19) for  $\mathcal{S}(\mathbf{p})$  given as (32). We recall the construction from [24, Lemma 2]. Define the matrix  $\mathbf{M}_0$  as follows:

$$\mathbf{M}_0 \stackrel{\text{def}}{=} \frac{\lambda^n}{\|\mathbf{h}_d\|_2^2} \begin{bmatrix} \lambda^n & \lambda^{n+1} & \lambda^{n+2} & \dots & \lambda^{2n-2} & -s \\ \lambda^{n+1} & \lambda^{n+2} & \dots & \lambda^{2n-2} & -s & 1 \\ \lambda^{n+2} & \dots & \lambda^{2n-2} & -s & 1 & \lambda \\ \vdots & \lambda^{2n-2} & -s & 1 & \dots & \vdots \\ \lambda^{2n-2} & -s & 1 & \dots & \dots & \lambda^{n-3} \\ -s & 1 & \lambda & \dots & \lambda^{n-3} & \lambda^{n-2} \end{bmatrix},$$

where  $s = \lambda \|\mathbf{h}_d\|_2^2$ . As shown in [24, Lemma 2],  $\|\mathbf{M}_0\|_2 = \lambda^n$ , and the matrix  $\mathbf{M}_0$  satisfies the equalities (25).

Since  $\|\mathbf{M}_0\|_2 < 1$ , we have that the conditions of Remark 16 are satisfied and thus the canonical completion  $\mathbf{p}_\lambda$  is the unique minimizer of (19). (Note that Theorem 1 only proves optimality of  $\mathbf{p}_\lambda$  for the problem (4).)

(ii) Here we again consider the case  $\lambda \in \mathbb{R}$ , but with  $\lambda \geq 1$ . Since we can scale the matrix without changing the optimality conditions (up to scaling), we can prove instead that for

$$\mathcal{F}(\mathbf{p}) = \begin{bmatrix} \lambda^{-d} & \dots & \lambda^{-1} & 1 \\ \vdots & \dots & \dots & p_1 \\ \lambda^{-1} & \dots & \dots & \vdots \\ 1 & p_1 & \dots & p_d \end{bmatrix},$$

the point  $\mathbf{p}_0 = [\lambda^{-1} \ \dots \ \lambda^{-d}]^\top$  is a minimizer of  $\|\mathcal{F}(\mathbf{p})\|_*$ . Define the flipping matrix as

$$\mathbf{J} = \begin{bmatrix} & & & 1 \\ & & & \\ & & & \\ 1 & & & \end{bmatrix}. \quad (34)$$

Then, by Lemma 22, for any  $\mathbf{p} \in \mathbb{C}^d$

$$\|\mathcal{F}(\mathbf{p})\|_* = \|\mathbf{J} \mathcal{F}(\mathbf{p})\|_* \geq \text{trace}\{\mathbf{J} \mathcal{F}(\mathbf{p})\} = n.$$

Next, we consider  $\mathbf{X} = \mathbf{J} \mathcal{F}(\mathbf{p}_0)$  and show that  $\mathbf{X}$  is positive semidefinite. For  $\lambda = 1$  this is obvious, and for  $\lambda > 1$ , as shown in [18, Example 1.2, page 27]

$$\mathbf{X}^{-1} = (1 - \lambda^{-2})^{-1} \begin{bmatrix} 1 & -\lambda^{-1} & 0 & \dots & 0 \\ -\lambda^{-1} & 1 + \lambda^{-2} & \dots & \dots & \vdots \\ 0 & \dots & \dots & \dots & 0 \\ \vdots & \dots & \dots & 1 + \lambda^{-2} & -\lambda^{-1} \\ 0 & \dots & 0 & -\lambda^{-1} & 1 \end{bmatrix}.$$

It is easy to see that  $(1 - \lambda^{-2})\mathbf{X}^{-1} = \mathbf{L}\mathbf{L}^\top$ , where

$$\mathbf{L} \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -\lambda^{-1} & 1 & \dots & \dots & \vdots \\ 0 & \dots & \dots & \dots & 0 \\ \vdots & \dots & \dots & 1 & 0 \\ 0 & \dots & 0 & -\lambda^{-1} & \sqrt{1 - \lambda^{-2}} \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Therefore,  $\mathbf{X}$  is positive semidefinite,  $\|\mathbf{X}\|_* = \text{trace}\{\mathbf{X}\} = n$ , and  $\mathbf{p}_0$  is a global minimizer of  $\|\mathcal{F}(\mathbf{p})\|_*$ .

(iii) The canonical completion after scaling (in terms of the matrix  $\mathcal{F}(\mathbf{p})$ ) corresponds to  $\mathbf{p}_1 = [\lambda \ \lambda^2 \ \dots \ \lambda^d]^\top$ .

By Lemma 22, we have that

$$\|\mathcal{F}(\mathbf{p}_1)\|_* \geq \text{trace}\{\mathcal{F}(\mathbf{p})\} = \sum_{j=0}^d \lambda^{2j-d} > d+1 = n,$$

where the last inequality holds since  $\lambda > 1$  and  $d > 1$ . Hence  $\mathbf{p}_1$  cannot be a minimizer.

(iv) The statements (ii) and (iii) are automatically valid for the real nuclear norm minimization problem (4). The same conclusion for (i) follows from Remark 15 or Theorem 1.

(v) Finally, we let  $\lambda \in \mathbb{C}$ . Then  $\lambda = \rho e^{i\theta}$ , where  $0 \leq \rho$  and  $\theta \in [0; 2\pi)$ . Now consider the matrix

$$\mathcal{G}(\mathbf{p}') = \begin{bmatrix} e^0 & & & \\ & e^{-i\theta} & & \\ & & \ddots & \\ & & & e^{-i\theta d} \end{bmatrix} \mathcal{S}(\mathbf{p}) \begin{bmatrix} e^0 & & & \\ & e^{-i\theta} & & \\ & & \ddots & \\ & & & e^{-i\theta d} \end{bmatrix},$$

where

$$p_k = e^{i\theta(d+k)} p'_k. \quad (35)$$

It is easy to see that the matrix is equal to

$$\mathcal{G}(\mathbf{p}') = \begin{bmatrix} 1 & \rho & \cdots & \rho^d \\ \rho & \rho^2 & \cdots & \rho^d \\ \vdots & \rho^2 & \cdots & \rho^d \\ \rho^d & \rho^d & \cdots & \rho^d \end{bmatrix},$$

which is of the form (32), and for which the statements (i), (ii), (iii) are proved.

The rest follows from the equality  $\|\mathcal{G}(\mathbf{p}')\|_* = \|\mathcal{S}(\mathbf{p})\|_*$ , invertibility of the transformation (35), and the fact that the completions  $p_{\lambda,k} = \lambda^{d+k}$  and  $p_{0,k} = \rho^{d-k} e^{i\theta(d+k)}$  are mapped to  $p'_{\rho,k} = \rho^{d+k}$  and  $p'_{0,k} = \rho^{d-k}$  respectively.  $\square$

### E. Proof of the main theorem

We now leave simple cases, and go back to the general formulation stated in Problem 1.

*Proof of Theorem 7.* (i) We only need to show that for any  $\varepsilon_{d,r}$  given in Lemma 21, there exists  $\rho_{d,r}$  such that for all  $\lambda_k$  satisfying the statements of the theorem it holds that

$$\|\mathbf{P} - \mathbf{P}_0\|_F \leq \varepsilon_{d,r},$$

where

$$\mathbf{P}_0 = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

The distance between projectors can be expressed as

$$\|\mathbf{P} - \mathbf{P}_0\|_F^2 = 2 \text{trace}\{\mathbf{P}_0(\mathbf{I} - \mathbf{P})\} = 2\|\mathbf{P}_0 \mathbf{U}_\perp\|_F^2,$$

where  $\mathbf{U}_\perp \in \mathbb{C}^{n \times (n-r)}$  is an orthonormal basis of the nullspace of  $\mathbf{P}$  (the reasoning is the same as in (29)).

The nullspace of  $\mathbf{P}$  coincides with the image of the matrix

$$\mathbf{K} = \begin{bmatrix} q_0 & 0 & \cdots & 0 \\ \vdots & q_0 & \ddots & \vdots \\ q_r & \vdots & \ddots & 0 \\ 0 & q_r & \ddots & q_0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & q_r \end{bmatrix} \in \mathbb{C}^{n \times (n-r)},$$

where  $q(z)$  is the characteristic polynomial (6) (with  $q_r = 1$ ). Then  $\mathbf{U}_\perp$  can be found from an SVD  $\mathbf{K} = \mathbf{U}_\perp \mathbf{\Sigma} \mathbf{V}_\perp^H$ , and, by submultiplicativity of matrix norms, we have that

$$\|\mathbf{P} - \mathbf{P}_0\|_F^2 \leq 2\|\mathbf{P}_0 \mathbf{K}\|_F^2 \|\mathbf{\Sigma}^{-1}\|_2^2$$

The non-zero entries of the matrix  $\mathbf{P}_0 \mathbf{K}$  are  $q_j$ , for  $j = \{0, \dots, r-1\}$ . By Vieta's formulae, these  $q_j$  are homogeneous polynomials in the roots  $\lambda_k$ . Therefore, there exist universal constants  $C_j$  (depending only on  $r$ ),  $j = 1, \dots, r$  such that  $(n-r)q_{r-j}^2 \leq C_j \rho^{2j}$ , where  $\rho \stackrel{\text{def}}{=} \max |\lambda_k|$ . Hence,

$$\|\mathbf{P}_0 \mathbf{K}\|_F^2 \leq \sum_{k=1}^r (n-r) q_{r-j}^2 \leq \rho^2 A(\rho),$$

where

$$A(\rho) \stackrel{\text{def}}{=} (C_1 + C_2 \rho^2 + \cdots + C_r \rho^{2(r-1)}).$$

Next, from the well-known bounds on eigenvalues of Toeplitz matrices [31], we know that if  $\rho \leq 1$

$$\sigma_{\min}^2(\mathbf{K}) \geq \min_{|z|=1, z \in \mathbb{C}} |q(z)|^2 \geq (1-\rho)^{2r}.$$

By combining it all together, we have

$$\|\mathbf{P} - \mathbf{P}_0\|_F^2 \leq \rho^2 \frac{2A(\rho)}{(1-\rho)^{2r}}. \quad (36)$$

Now, we fix an arbitrary  $0 < \rho_0 < 1$  (say  $\rho_0 = 0.5$ ). For  $0 < \rho \leq \rho_0$ , it follows from (36) that

$$\|\mathbf{P} - \mathbf{P}_0\|_F^2 \leq \rho^2 C_{\rho_0, r},$$

where  $C_{\rho_0, r}$  (that depends only on  $\rho_0$  and  $r$ ) is defined as

$$C_{\rho_0, r} \stackrel{\text{def}}{=} \frac{2A(\rho_0)}{(1-\rho_0)^{2r}}.$$

Finally, we define the constant of Theorem 7 to be

$$\rho_{d,r} \stackrel{\text{def}}{=} \min \left( \rho_0, \frac{\varepsilon_{d,r}}{\sqrt{C_{\rho_0, r}}} \right), \quad (37)$$

for which  $\|\mathbf{P} - \mathbf{P}_0\|_F \leq \varepsilon_{d,r}$  if  $\rho \leq \rho_{d,r}$ .

(ii) This part is a direct consequence of Remark 15.  $\square$

## V. NUMERICAL RESULTS

Theorem 7 mainly states the existence of such radius  $\rho_{d,r}$ . But the radius that can be deduced from the proofs can be very small. For example, from (30) it is easy to show that  $\varepsilon_{d,r} = \frac{1}{2d}$  can be chosen in Lemma 21. Then we can get  $\rho_{d,r}$  from (37), which would give a very conservative lower bound on possible  $\rho_{d,r}$ . This is the reason why we report numerical experiments aiming at finding the largest lower bound in this section.

The MATLAB package CVXOPT [32] (with default settings<sup>3</sup>) is used for nuclear norm minimization. The MATLAB code that reproduces the results will be made available together with the authors' version of the paper at <https://hal.archives-ouvertes.fr/hal-01130631>.

<sup>3</sup>For solving the complex nuclear norm minimization problem (3), we used an equivalent real nuclear norm minimization problem (see (38) and (39) in Appendix A). At the moment of writing this article, the direct minimization of the nuclear norm of a complex matrix in CVXOPT gave a poor performance.

### A. The case of 1 or 2 exponents

The setup for the following experiments will be similar. We take a specific  $\mathbf{h}_d \in \mathbb{C}^{d+1}$ , find the solution of (2) (or the solution of (4) if  $\mathbf{h}_d \in \mathbb{R}^{d+1}$ ), and measure the Frobenius norm between the solution and the canonical completion (10).

1) *The rank-one case:* First, we consider the rank-one case. We take  $n = 6$ , and  $h_k = \lambda^k$ , where  $\lambda = a + bi$ ,  $a, b \in (-1, 1)^2$ , and plot the results in Fig. 1.

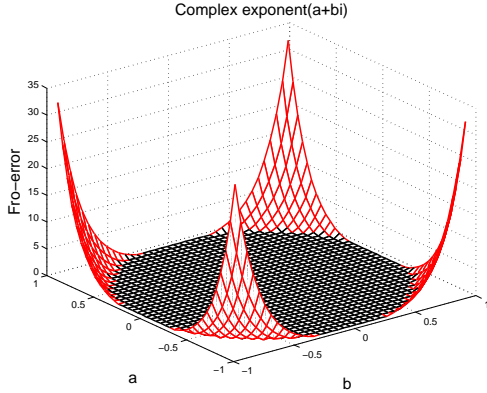


Fig. 1. Nuclear norm reconstruction (Frobenius distance),  $h_k = \lambda^k$ ,  $\lambda = a + bi$ . Black area correspond to the values less than  $10^{-6}$ .

The results in Fig. 1 confirm the results of Theorem 6: the canonical completion for  $|\lambda| > 1$  is not recovered by the nuclear norm relaxation.

2) *The rank-two case:* We consider real sequences  $\mathbf{h}_d$  with  $\text{hrank}(\mathbf{h}_d) = 2$ . The following three situations are possible:

- 1)  $\lambda_1, \lambda_2 \in \mathbb{R}$  (two simple real roots);
- 2)  $\lambda_2 = \overline{\lambda_1} \notin \mathbb{R}$  (two simple complex conjugate roots);
- 3)  $\lambda_1 \in \mathbb{R}, \nu_1 = 2$  (double real root).

The first case is considered in [24, Fig. 1], where it is shown (numerically) that the radius  $\rho$  is less than 1. In this section, we examine the second and the third cases. We generate the corresponding  $\mathbf{h}_d$  (in the second case  $c_1 = c_2$ ) and compute the Frobenius distance between the solutions of (2) and (4).

In Fig. 2, we plot the nuclear norm reconstruction errors for a  $6 \times 6$  matrix and the last two cases. As seen in Fig. 2, the limiting modulus is also strictly less than 1. The radius is smaller in the case of a double root and also in the case when two conjugate roots are close to each other.

### B. The case of multiple exponents

In this experiment, we aim at estimating the radius  $\rho$  based on random realizations of  $\lambda_k$ . We fix  $n = 9$ , and for each  $r \in \{1, \dots, 4\}$  and for  $\rho \in (0, 1)$ , we generate  $\{\lambda_1, \dots, \lambda_r\}$ , such that  $|\lambda_1| = \rho$  and  $|\lambda_k| \leq \rho$ . All the  $c_k$  are equal to 1.

We consider two situations (real and complex roots):

- $\lambda_k = \rho_k$  (in this case  $\rho_1 = \rho$ , and  $\rho_k, k > 1$  are independent and uniformly distributed in  $[\rho; -\rho]$ )
- $\lambda_k = \rho_k e^{i\pi\phi_k}$ , where  $\rho_k$  are as in the previous case,  $\phi_k$  are independent, uniformly distributed in  $[0; 1]$ , and independent of  $\rho_k$ .

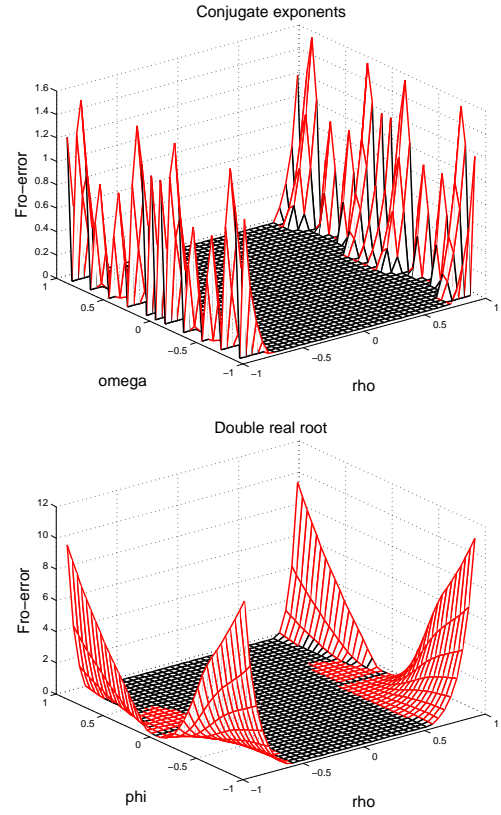


Fig. 2. Nuclear norm reconstruction (Frobenius distance). Top: for two complex conjugate roots  $\lambda_{1,2} = \rho \cdot \exp(\pm i\pi\omega)$  (corresponds to the damped cosine sequence  $h_t = \rho^t \cdot \cos(\pi(\omega + 1)t)$ ). Bottom: for a double root  $\rho$  (corresponds to the damped linear function  $h_t = (t \cdot \tan(0.75 \cdot \pi \cdot \varphi) + 1) \cdot \rho^t$ ). Black areas correspond to the values less than  $10^{-6}$ .

We repeat the experiment  $M = 100$  times, and select the maximum Frobenius error across all the realizations. The results plotted in Fig. 3 confirm the conclusions of Theorem 7. We see that the nuclear norm heuristic works up to a certain modulus  $\rho$ . Note that the bounds on  $\rho$  are similar for the real and complex cases.

## APPENDIX A

*Proof of Proposition 14.* In order to derive the optimality conditions for the complex-valued case, we construct an extended real-valued map  $\mathcal{S}_{ext} : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2n \times 2n}$  as follows. For  $\mathbf{p}_{ext} = [\mathbf{p}_R^T \ \mathbf{p}_I^T]$ , where  $\mathbf{p}_R, \mathbf{p}_I \in \mathbb{R}^N$ , we define

$$\mathcal{S}_{ext}(\mathbf{p}_{ext}) \stackrel{\text{def}}{=} \begin{bmatrix} \mathcal{S}(\mathbf{p}_R) & -\mathcal{S}(\mathbf{p}_I) \\ \mathcal{S}(\mathbf{p}_I) & \mathcal{S}(\mathbf{p}_R) \end{bmatrix} \quad (38)$$

From [33], the singular values of  $\mathcal{S}_{ext}(\mathbf{p}_{ext})$  contain two copies of singular values of  $\mathcal{S}(\mathbf{p})$  (with  $\mathbf{p} \stackrel{\text{def}}{=} \mathbf{p}_R + i\mathbf{p}_I$ ). Hence, the problem (19) is equivalent to the real nuclear norm minimization problem<sup>4</sup>

$$\hat{\mathbf{p}}_{ext} = \arg \min_{\mathbf{p}_{ext} \in \mathbb{R}^{2N}} \|\mathcal{S}_{ext}(\mathbf{p}_{ext})\|_*. \quad (39)$$

Consider an SVD of the complex symmetric matrix  $\mathcal{S}(\mathbf{p})$ ,

$$\mathcal{S}(\mathbf{p}) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \quad (40)$$

<sup>4</sup>In fact, the corresponding rank minimization problems are also equivalent.

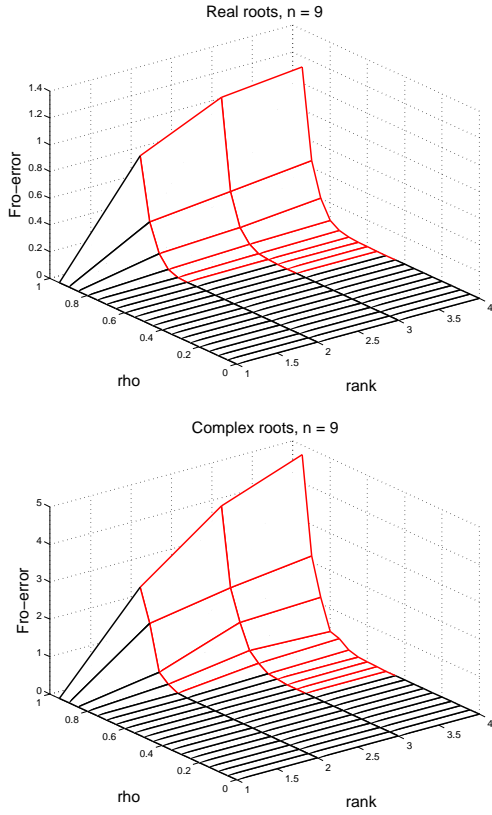


Fig. 3. Nuclear norm reconstruction,  $n = 9$ ,  $M = 100$  (Frobenius distance). Top: real roots, bottom: complex roots. Black area corresponds to the values less than  $10^{-5}$ .

where  $U^H U = V^H V = I_r$ , and  $\Sigma \in \mathbb{R}^{r \times r}$  is diagonal. Since  $\mathcal{S}(p)$  is symmetric,  $\mathcal{S}(p) = \bar{V} \Sigma U^T$  is also an SVD.

Next, we have that the matrices  $B = B_{\mathcal{R}} + iB_{\mathcal{I}}$  and  $P = P_{\mathcal{R}} + iP_{\mathcal{I}}$  have the form

$$\begin{aligned} B_{\mathcal{R}} &= U_{\mathcal{R}} V_{\mathcal{R}}^T + U_{\mathcal{I}} V_{\mathcal{I}}^T, & B_{\mathcal{I}} &= U_{\mathcal{I}} V_{\mathcal{R}}^T - U_{\mathcal{R}} V_{\mathcal{I}}^T, \\ P_{\mathcal{R}} &= U_{\mathcal{R}} U_{\mathcal{R}}^T + U_{\mathcal{I}} U_{\mathcal{I}}^T = V_{\mathcal{R}} V_{\mathcal{R}}^T + V_{\mathcal{I}} V_{\mathcal{I}}^T, \\ P_{\mathcal{I}} &= U_{\mathcal{I}} U_{\mathcal{R}}^T - U_{\mathcal{R}} U_{\mathcal{I}}^T = V_{\mathcal{R}} V_{\mathcal{I}}^T - V_{\mathcal{I}} V_{\mathcal{R}}^T. \end{aligned}$$

Since the matrix  $P$  is Hermitian and  $B$  is complex symmetric, we have  $P_{\mathcal{R}}^T = P_{\mathcal{R}}$ ,  $P_{\mathcal{I}}^T = -P_{\mathcal{I}}$ ,  $B_{\mathcal{R}}^T = B_{\mathcal{R}}$  and  $B_{\mathcal{I}}^T = B_{\mathcal{I}}$ . From [33], the matrix  $\mathcal{S}_{ext}(p_{ext})$  admits the SVD

$$\mathcal{S}_{ext}(p_{ext}) = U_{ext} \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \Sigma \end{bmatrix} V_{ext}^T, \quad (41)$$

where for  $U = U_{\mathcal{R}} + iU_{\mathcal{I}}$  and  $V = V_{\mathcal{R}} + iV_{\mathcal{I}}$

$$U_{ext} \stackrel{\text{def}}{=} \begin{bmatrix} U_{\mathcal{R}} & -U_{\mathcal{I}} \\ U_{\mathcal{I}} & U_{\mathcal{R}} \end{bmatrix}, \quad V_{ext} \stackrel{\text{def}}{=} \begin{bmatrix} V_{\mathcal{R}} & -V_{\mathcal{I}} \\ V_{\mathcal{I}} & V_{\mathcal{R}} \end{bmatrix}.$$

We can express the matrix  $B_{ext}$  for the structure  $\mathcal{S}_{ext}$  as

$$B_{ext} \stackrel{\text{def}}{=} U_{ext} V_{ext}^T = \begin{bmatrix} B_{\mathcal{R}} & -B_{\mathcal{I}} \\ B_{\mathcal{I}} & B_{\mathcal{R}} \end{bmatrix}.$$

Next, from (38), we have that

$$\begin{aligned} \mathcal{S}_{ext}(p_{ext}) &= \begin{bmatrix} S_{0,\mathcal{R}} & -S_{0,\mathcal{I}} \\ S_{0,\mathcal{I}} & S_{0,\mathcal{R}} \end{bmatrix} + \\ & \sum_{k=1}^N p_{k,\mathcal{R}} \begin{bmatrix} S_k & \mathbf{0} \\ \mathbf{0} & S_k \end{bmatrix} + \sum_{k=1}^N p_{k,\mathcal{I}} \begin{bmatrix} \mathbf{0} & -S_k \\ S_k & \mathbf{0} \end{bmatrix}, \end{aligned} \quad (42)$$

where  $S_0 = S_{0,\mathcal{R}} + iS_{0,\mathcal{I}}$ . In what follows, we prove statements of the proposition.

(i) From (17) and (18), the subgradient of  $g(p_{ext}) \stackrel{\text{def}}{=} \|\mathcal{S}_{ext}(p_{ext})\|_*$  is equal to

$$\left\{ \begin{bmatrix} a(M_{ext}) \\ b(M_{ext}) \end{bmatrix} : M_{ext} \in \mathbb{R}^{2N \times 2N}, \|M_{ext}\|_2 \leq 1 \right\},$$

where

$$\begin{aligned} a_k(M_{ext}) &= \left\langle \begin{bmatrix} S_k & \mathbf{0} \\ \mathbf{0} & S_k \end{bmatrix}, B_{ext} + Q_{ext} M_{ext} Q_{ext}^T \right\rangle_F, \\ b_k(M_{ext}) &= \left\langle \begin{bmatrix} \mathbf{0} & -S_k \\ S_k & \mathbf{0} \end{bmatrix}, B_{ext} + Q_{ext} M_{ext} Q_{ext}^T \right\rangle_F, \end{aligned}$$

and for  $Q = Q_{\mathcal{R}} + iQ_{\mathcal{I}} \stackrel{\text{def}}{=} I_n - P$

$$Q_{ext} = \begin{bmatrix} Q_{\mathcal{R}} & -Q_{\mathcal{I}} \\ Q_{\mathcal{I}} & Q_{\mathcal{R}} \end{bmatrix}, \quad Q_{ext}^T = \begin{bmatrix} Q_{\mathcal{R}} & Q_{\mathcal{I}} \\ -Q_{\mathcal{I}} & Q_{\mathcal{R}} \end{bmatrix}$$

are the projectors on the left and right nullspaces of  $\mathcal{S}_{ext}(p_{ext})$ . Immediately, we have that

$$\begin{aligned} 2\langle S_k, P \rangle_F &= \left\langle \begin{bmatrix} S_k & \mathbf{0} \\ \mathbf{0} & S_k \end{bmatrix}, B_{ext} \right\rangle_F \\ &+ i \left\langle \begin{bmatrix} \mathbf{0} & -S_k \\ S_k & \mathbf{0} \end{bmatrix}, B_{ext} \right\rangle_F. \end{aligned}$$

Now let us define the matrices

$$\begin{aligned} M_{ext} &= \begin{bmatrix} M_1 & -M_3 \\ M_2 & M_4 \end{bmatrix}, \quad M_{\mathcal{R}} \stackrel{\text{def}}{=} \frac{M_1 + M_4}{2}, \\ M_{\mathcal{I}} &\stackrel{\text{def}}{=} \frac{M_2 + M_3}{2}, \quad M \stackrel{\text{def}}{=} M_{\mathcal{R}} + iM_{\mathcal{I}}. \end{aligned} \quad (43)$$

Easy calculations show that the matrix  $Q_{ext} M_{ext} Q_{ext}^T$  can be expressed as

$$Q_{ext} M_{ext} Q_{ext}^T = \begin{bmatrix} C_1 & -C_3 \\ C_2 & C_4 \end{bmatrix}, \quad (44)$$

where

$$\begin{aligned} C_1 &= Q_{\mathcal{R}} M_1 Q_{\mathcal{R}} + Q_{\mathcal{R}} M_3 Q_{\mathcal{I}} - Q_{\mathcal{I}} M_2 Q_{\mathcal{R}} + Q_{\mathcal{I}} M_4 Q_{\mathcal{I}}, \\ C_2 &= Q_{\mathcal{R}} M_2 Q_{\mathcal{R}} - Q_{\mathcal{R}} M_4 Q_{\mathcal{I}} + Q_{\mathcal{I}} M_1 Q_{\mathcal{R}} + Q_{\mathcal{I}} M_3 Q_{\mathcal{I}}, \\ C_3 &= Q_{\mathcal{R}} M_3 Q_{\mathcal{R}} - Q_{\mathcal{R}} M_1 Q_{\mathcal{I}} + Q_{\mathcal{I}} M_4 Q_{\mathcal{R}} + Q_{\mathcal{I}} M_2 Q_{\mathcal{I}}, \\ C_4 &= Q_{\mathcal{R}} M_4 Q_{\mathcal{R}} + Q_{\mathcal{R}} M_2 Q_{\mathcal{I}} - Q_{\mathcal{I}} M_3 Q_{\mathcal{R}} + Q_{\mathcal{I}} M_1 Q_{\mathcal{I}}, \end{aligned}$$

and

$$QM Q^T = \frac{C_1 + C_4}{2} + i \frac{C_2 + C_3}{2}.$$

Then, immediately,

$$\begin{aligned} &2\langle S_k, B + QM Q^T \rangle_F \\ &= \left\langle \begin{bmatrix} S_k & \mathbf{0} \\ \mathbf{0} & S_k \end{bmatrix}, B_{ext} - Q_{ext} M_{ext} Q_{ext}^T \right\rangle_F \\ &+ i \left\langle \begin{bmatrix} \mathbf{0} & -S_k \\ S_k & \mathbf{0} \end{bmatrix}, B_{ext} - Q_{ext} M_{ext} Q_{ext}^T \right\rangle_F. \end{aligned}$$

Hence, we obtain that

$$a_k(M_{ext}) + i b_k(M_{ext}) = \langle S_k, B + QM Q^T \rangle_F$$

Finally, it is easy to see that for  $M$  defined as in (43)

$$\begin{aligned} \|M\|_2 &= \left\| \begin{bmatrix} M_{\mathcal{R}} & -M_{\mathcal{I}} \\ M_{\mathcal{I}} & M_{\mathcal{R}} \end{bmatrix} \right\|_2 \\ &= \left\| \frac{M_{ext} + \begin{bmatrix} \mathbf{0} & -I \\ I & \mathbf{0} \end{bmatrix} M_{ext} \begin{bmatrix} \mathbf{0} & -I \\ I & \mathbf{0} \end{bmatrix}}{2} \right\|_2 \\ &\leq \|M_{ext}\|_2, \end{aligned} \quad (45)$$

where the equality takes place if  $M_1 = M_4$  and  $M_2 = M_3$ . Hence, if  $\|M_{ext}\|_2 < 1$  and then  $\|M\|_2 < 1$ , according to (45). Vice versa, if we are given  $M = M_{\mathcal{R}} + iM_{\mathcal{I}}$  with  $\|M\|_2 < 1$ , by taking  $M_1 = M_4 = M_{\mathcal{R}}$  and  $M_2 = M_3 = M_{\mathcal{I}}$ , we obtain  $M_{ext}$  with  $\|M_{ext}\|_2 = \|M\|_2$ , which completes the proof.

(ii) Let  $M_{ext}$  be as in (43), with  $M_1 = M_4 = M_{\mathcal{R}}$ ,  $M_2 = M_3 = M_{\mathcal{I}}$ . Then, in view of (44), we have that

$$QM\mathbf{Q}^T = \mathbf{0} \iff Q_{ext}M_{ext}Q_{ext}^T = \mathbf{0},$$

which completes the proof.  $\square$

#### ACKNOWLEDGMENT

This work is supported by the ERC project ‘‘DECODA’’ no.320594, in the frame of the European program FP7/2007-2013. We thank L. Dai and K. Pelckmans for motivating discussions. We also thank the anonymous reviewers for their useful remarks that led to an improved presentation and a simpler proof of Lemma 22.

#### REFERENCES

- [1] Y. Chen and Y. Chi, ‘‘Robust spectral compressed sensing via structured matrix completion,’’ *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6576–6601, Oct 2014.
- [2] E. J. Candès and B. Recht, ‘‘Exact matrix completion via convex optimization,’’ *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [3] D. Gross, ‘‘Recovering low-rank matrices from few coefficients in any basis,’’ *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.
- [4] B. De Schutter, ‘‘Minimal state-space realization in linear system theory: an overview,’’ *J. Comput. Appl. Math.*, vol. 121, no. 1, pp. 331–354, 2000.
- [5] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, ‘‘Hankel matrix rank minimization with applications in system identification and realization,’’ *SIAM J. Matrix Ana. Appl.*, vol. 34, no. 3, pp. 946–977, 2013.
- [6] Z. Liu and L. Vandenberghe, ‘‘Interior-point method for nuclear norm approximation with application to system identification,’’ *SIAM J. Matrix Ana. Appl.*, vol. 31, no. 3, pp. 1235–1256, 2009.
- [7] M. W. Trosset, ‘‘Distance matrix completion by numerical optimization,’’ *Comput. Opt. Appl.*, vol. 17, no. 1, pp. 11–22, 2000.
- [8] R. M. Gray and L. D. Davisson, *An Introduction to Statistical Signal Processing*. London: Cambridge University Press, 2005.
- [9] R. Bro, ‘‘Parafac, tutorial and applications,’’ *Chemom. Intel. Lab. Syst.*, vol. 38, pp. 149–171, 1997.
- [10] L. Hogben, ‘‘Matrix completion problems for pairs of related classes of matrices,’’ *Lin. Alg. Appl.*, vol. 373, pp. 13–29, 2003.
- [11] N. Kreimer, A. Stanton, and M. D. Sacchi, ‘‘Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction,’’ *Geophysics*, vol. 78, no. 6, pp. V273–V284, 2013.
- [12] E. J. Candès and Y. Plan, ‘‘Matrix completion with noise,’’ *Proc. of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.
- [13] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas, ‘‘Symmetric tensor decomposition,’’ *Linear Algebra and its Applications*, vol. 433, no. 11–12, pp. 1851–1872, 2010.
- [14] M. Fazel, ‘‘Matrix rank minimization with applications,’’ Ph.D. dissertation, Elec. Eng. Dept, Stanford University, 2002.
- [15] N. Gillis and F. Glineur, ‘‘Low-rank matrix approximation with weights or missing data is NP-hard,’’ *SIAM Journal on Matrix Analysis and Applications*, vol. 32, no. 4, pp. 1149–1165, 2011.
- [16] H. Woerdeman, ‘‘Minimal rank completions for block matrices,’’ *Linear Algebra and its Applications*, vol. 121, pp. 105–122, 1989.
- [17] I. Iohvidov, *Toeplitz and Hankel matrices and forms. Algebraic theory*. Birkhäuser, Basel, 1982.
- [18] G. Heinig and K. Rost, *Algebraic methods for Toeplitz-like matrices and operators*. Birkhäuser, Boston, 1984.
- [19] S. Feldmann and G. Heinig, ‘‘Parametrization of minimal rank block Hankel matrix extensions and minimal partial realizations,’’ *Integral Equations and Operator Theory*, vol. 33, no. 2, pp. 153–171, 1999.
- [20] M. Laurent and B. Mourrain, ‘‘A generalized flat extension theorem for moment matrices,’’ *Archiv der Mathematik*, vol. 93, no. 1, pp. 87–98, 2009.
- [21] B. Recht, M. Fazel, and P. Parrilo, ‘‘Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,’’ *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [22] I. Markovsky and K. Usevich, ‘‘Structured low-rank approximation with missing data,’’ *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 2, pp. 814–830, 2013.
- [23] 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.
- [24] L. Dai and K. Pelckmans, ‘‘On the nuclear norm heuristic for a Hankel matrix completion problem,’’ *Automatica*, vol. 51, pp. 268–272, 2015.
- [25] H. Butcher and J. Gillard, ‘‘Simple nuclear norm based algorithms for imputing missing data and forecasting in time series,’’ *Statistics and Its Interface*, 2016, to appear.
- [26] O. Bosgra, ‘‘On parametrizations for the minimal partial realization problem,’’ *Systems & Control Letters*, vol. 3, no. 4, pp. 181–187, 1983.
- [27] I. Markovsky, ‘‘On the most powerful unfalsified model for data with missing values,’’ *Systems & Control Letters*, 2016.
- [28] K. Usevich and P. Comon, *Quasi-Hankel Low-Rank Matrix Completion: a Convex Relaxation*, 2015, preprint, <http://arxiv.org/abs/1505.07766>.
- [29] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer Verlag, Heidelberg, 2001.
- [30] G. Watson, ‘‘Characterization of the subdifferential of some matrix norms,’’ *Linear Algebra and its Applications*, vol. 170, pp. 33–45, 1992.
- [31] U. Grenander and G. Szegő, *Toeplitz forms and their applications*, 2nd ed. Chelsea, New York, 1984.
- [32] M. Grant and S. Boyd, ‘‘CVX: Matlab software for disciplined convex programming, version 2.1,’’ <http://cvxr.com/cvx>, Mar. 2014.
- [33] D. Day and M. Heroux, ‘‘Solving complex-valued linear systems via equivalent real formulations,’’ *SIAM Journal on Scientific Computing*, vol. 23, no. 2, pp. 480–498, 2001.



**Konstantin Usevich** obtained a specialist degree in 2007 and Ph.D. degree in 2011, both from the Department of Statistical Modeling, St.Petersburg State University, Russia. He held postdoctoral research positions at the University of Southampton, UK and at the Department ELEC, Vrije Universiteit Brussel, Belgium. Since July 2014, he is a postdoctoral researcher at GIPSA-lab, Centre National de la Recherche Scientifique, France. His research interests are in low-rank approximations of structured matrices and tensors, signal and image processing, system identification, and approximate polynomial computations.



**Pierre Comon** (M87 SM95 F07) graduated in 1982, and received the Doctorate degree in 1985, both from the University of Grenoble, France. He later received the Habilitation to Lead Researches in 1995, from the University of Nice, France. He has been for nearly 13 years in industry, first with Crouzet-Sextant, Valence, France, between 1982 and 1985, and then with Thomson Marconi, Sophia Antipolis, France, between 1988 and 1997. He spent 1987 with the ISL laboratory, Stanford University, CA. He joined in 1997 the Eurecom Institute, Sophia Antipolis, France. He is research director with CNRS since 1998, first at laboratory I3S, Sophia Antipolis, France, until 2012, and then at Gipsa-Lab, Grenoble, France. His research interests include High-Order Statistics (HOS), Blind techniques, Statistical Signal and Array Processing, Tensor decompositions, Multi-Way Factor Analysis, Data Mining and its applications to biomedical end environment.

Dr. Comon was Associate Editor of the *IEEE Transactions on Signal Processing* from 1995 to 1998, and a member of the French National Committee of Scientific Research from 1995 to 2000. He was the coordinator of the European Basic Research Working Group on HOS, *ATHOS*, from 1992 to 1995. Between 1992 and 1998, he was a member of the Technical and Scientific Council of the Thomson Group. Between 2001 and 2004 he acted as launching Associate Editor with the *IEEE Transactions on Circuits and Systems I*, in the area of Blind Techniques. He has also been a member of the editorial board of the Elsevier journal *Signal Processing* from 2006 to 2011, and member of several IEEE TC. He is presently in the editorial board of *SIAM Journal on Matrix Analysis and Applications*.